UNIVERSITY OF LIVERPOOL

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Conor Langmuir Proud

Abstract

Examining the Role of Transformation in the Acquisition of Resistance and Rescue of Genomes from Oxidative Stress

Conor Langmuir Proud

The pneumococcus remains a pathogen of global importance due to its ability to circumvent antibiotic and vaccine pressures. For resource poor countries such as Malawi, the threat posed by this pathogen is perhaps the strongest, as access to timely and effective healthcare is limited. Coupled with this, high rates of HIV mean that pneumococcal disease is particularly prevalent in Malawi. The recent rollout of vaccination, anti-retroviral therapy to combat HIV, and increased availability to antimicrobials is helping to control this pathogen. However, recent genetic analyses have already demonstrated the potential for the pneumococcus to circumvent such measures, through vaccine-escape, and antibiotic resistance.

The pneumococcus inhabits the nasopharynx where it is thought to recombine with other closely related mitis group species. Such recombination events are thought to have allowed this pathogen to first acquire beta-lactam resistance- a class of antibiotics favoured for the treatment of pneumococcal diseases. Survival within this niche is nonetheless challenging, as it is subject to high levels of attack by charged oxygen particles, termed oxidative stress. Oxidative stress results not only from attack by the host's immune system, but also from the pneumococci's metabolism. Such attack has been found to cause a wide variety of genetic lesions in the pneumococcus, and consequently the repair of such damage is likely to be essential for survival. Despite this, the pneumococcus lacks many of the normal repair machineries found in bacteria such as *E. coli*, and as such some have suggested a role for recombination in this repair process.

Detection of genetic damage within collections of clinical pneumococcal isolates *in silico* indicated deletional damage occurred frequently, and genome-wide. This

implied that the damage detected was not a consequence of replication errors, which would be expected to occur clustered close to the origin of replication. Deletion repair was identified as an essential mechanism in order to subjugate the fitness costs associated with such damage *in situ.* Recombination with siblings (self-repair) appeared to be the dominant force in this process. Attempts to model genetic damage *in vitro* were however limited, with exogenously delivered hydrogen peroxide appearing to poorly reflect the natural condition. It may therefore be important to consider pneumococci subject to abnormal levels of oxidative stress in clinical settings to better characterise this repair process.

Recombination has additionally played a major role in the ability for pneumococci to respond to clinical interventions. The ability for pneumococci, and other mitis group species to circulate globally was found to play an important role in the acquisition of beta-lactam resistance in the Malawian pneumococcal population. Continued epidemiological surveillance of this pathogen will therefore be important in order to assess how this pathogen responds to the current clinical interventions, such as PCV13 vaccination. Although recombination within the mitis group was identified, the interactions between pneumococci and other members of this group remain poorly characterised. Such events however appear infrequent, likely restricted to a subset of the population, such as in infants.

Finally, to better elucidate the pathways by which beta-lactam resistance arises clinically; a genome wide approach was taken to identifying resistance associated genes. In addition to indicating novel resistance mechanisms, a role for interspecies recombination within these sites was identified.

Acknowledgements

This PhD has been carried out across multiple countries and research institutions, and as such I am thankful to the multitude of people who have either helped directly by offering research advice, or indirectly by making my stay across these different locations so enjoyable.

Dean Everett has offered enthusiasm and support throughout the period of this PhD, in addition to helping me develop as a researcher. Stephen Bentley has similarly been forthcoming with advice and has ensured I have been well looked after during my time spent at the Sanger Institute. Finally, Chris Dowson has helped me develop ideas, and provided a great deal of insight into pneumococcal biology. As such, the contribution of all three supervisors has been invaluable.

At the Sanger Institute Simon Harris has provided much the technical knowledge that allowed me to undertake the bioinformatics work covered. Susannah Salter has similarly offered great laboratory advice and help when required.

At the University of Warwick I am grateful to Edward Sullivan for his help and instructions on carrying out the sorbarod work. In addition, I am thankful to Karen Hinxman for providing me with a place to live when I first arrived. My friendships with Richard Edwards, Lucy Wright and Jure Vajs have similarly ensured that my time at Warwick was an enjoyable one. Finally, I am grateful to Smita Chauhan for her friendship and advice, which have helped me cope with the pressures of completing the projects discussed herein.

Finally, my time spent in Malawi has given me memories and experiences for life. It gave me an insight into the numerous challenges faced by such impoverished nations, and the difficulties posed to those dedicated to researching and improving the medical care in these countries. From the research group in Malawi I am particularly thankful to Maaike Alaerts, who helped with my lab work, and Chrispin Chaguza- who managed to teach me how to learn Python in an hour, which has been an essential tool during my research. Louise Afran has

similarly been a great friend to me in Malawi and since, without whom my time in Malawi I am sure would have been far less enjoyable and memorable.

Finally, I'd like to thank all those others who have helped in their different ways over the last fours years I have spent researching the pneumococcus.

Dedication

To my family and friends, for their encouragement, support and inspiration.

Table of Contents

List of Tables

List of Figures

Abbreviations

Acquired Immunodeficiency Syndrome	AIDS
Adenosine Tri Phosphate	ATP
Amino Acid	AA
Accelerated Development and Introduction Plan	PneumoADIP
Basic Local Alignment Search Tool	BLAST
Binary Alignment Map	BAM
Brain Heart Infusion	BHI
British Society of Antimicrobial Chemotherapy	BSAC
Centre for Disease Control and Prevention	CDC
Choline Binding Protein	CBP
Clonal Complex	CC
Coding DNA Sequence	CDS
Competence Stimulating Peptide	CSP
Dimethyl Sulphoxide	DMSO
dNTP	Deoxynucleotide triphosphate
European Nucleotide Archive	ENA
Double Strand Break	DSB
Double Strand DNA	dsDNA
Global Alliance for Vaccines and Immunisation	GAVI
Hepatitis influenzae B	HiB
Human Immunodeficiency Virus	HIV
Hydrogen Peroxide	H2O2
Immunoglobulin	Ig
Integrative Conjugative Element	ICE
Interleukin	IL
Invasive Pneumococcal Disease	IPD
Iso-sensitest	IS
Low Molecular Mass	LMM
Malawi-Liverpool Wellcome Trust	MLW
Minimum Inhibitory Concentration	MIC
Mobile Genetic Element	MGE
Multi Drug Resistance	MDR
Multi Locus Sequence Type	MLST
N-acetylglucosamine	GlcNAc
N-acetylmuramic acid	MurNAc
N-acetylmuramylpentapeptide	MurNAc pentapeptide
Next Generation Sequencing	NGS
PCR	polymerase chain reaction
Penicillin Binding Protein	PBP

Penicillin-binding protein And Serine/Threonine kinase Associated domains	PASTA
Phosphoglucomutase	PGM
Phosphate Buffered Saline	PBS
Platelet Activating Factor receptor	PAFr
Pneumococcal Collagen Like Protein A	PclA
Pneumococcal Conjugate Vaccine	PCV
Pneumococcal Molecular Epidemiology Network	PMEN
Pneumococcal Polysaccharide Vaccine	PPV
Pneumococcal Serine-Rich Repeat Protein	psrP
Pneumococcal vaccines Accelerated Development and Introduction Plan	PneumoADIP
Pneumolysin	Ply
Polymeric immunoglobulin receptor	plgR
Pulsed Field Gel Electophoresis	PFGE
Pyruvate oxidase	SpxB
Quantitative Biology Centre	QuBIC
RNAP	RNA polymerase
Second generation sequencing technology	SGST
Single-Nucleotide Polymorphism	SNP
Single Strand Break	SSB
Single Strand DNA	ssDNA
Skim milk, tryptone, glucose, and glycerine	STGG
Tricarboxylic acid cycle	TCA
Tumour Necrosis Factor	TNF
Ultraviolet	UV
Universal Transport Medium	UTM
University of Liverpool	UoL
University of Warwick	UoW
Wellcome Trust Sanger Institute	WTSI
Whole Genome Sequence	WGS
World Health Organisation	WHO

1 Introduction

1.1 A History of the Pneumococcus

The pneumococcus is estimated to cause 1.6 million deaths annually, with up to 1 million of these occurring in children (<5 years) (WHO, 2015). In 1918 William Osler famously branded the pneumococcus "Captain of the men of death", and accounts of pneumococcal diseases precede the Ancient Greeks (White et al., 2009, Singer, 1928). It is frequently recognised as the leading cause of pneumonia (Fang et al., 1990, Welte, 2012, Johansson et al., 2014, Walker et al., 2013), while remaining a prominent cause of otitis media (Bluestone et al., 1992, Klein, 1994, Sierra et al., 2011), bacteraemia (Plouffe et al., 1996, Gordon et al., 2001, Pilishvili et al., 2010) and meningitis (Durand et al., 1993, Schuchat et al., 1997, O'Brien et al., 2009).

1.1.1 Discovery

Kleb (1875), Eberth (1880), and Mátray (1880) offered a number of early descriptions of this pathogenic agent (White, 1938). However, they failed to carry out the experimental assays necessary to confirm the ability of this microorganism to cause disease. As such, it is the names of Sternberg, and Pasteur that have became synonymous with the initial discovery of the pneumococcus (White, 1938). In 1881, US Army surgeon George Sternberg (Sternberg, 1881) and Louise Pasteur, France, (Pasteur, 1881) both independently isolated pneumococci from the blood following subcutaneous inoculation of rabbits with human saliva. After inoculating a rabbit with his own saliva, Sternberg noted how a septicaemia was "produced by the rapid multiplication in the body of the infected animal, of a parasitic organism" (Sternberg, 1881, Sternberg, 1982). Whilst the pathogenic potential of this coccus had been demonstrated, in both cases, pneumococci had been isolated from asymptomatic carriage- the relationship between this coccus and the human diseases it caused consequently remained unknown.

Attempting to identify the cause of lobar pneumonia, Friedlander isolated these cocci from alveolar exudates (1883) and later from the blood of pneumonic patients (1884). However, his later attribution of lobar pneumonia to a bacillus meant that he is now better remembered for this pioneering use of the Gram stain for bacterial identification (Gram, 1884). Fraenkel (1984), using an animal model, would later attribute lobar pneumonia to a coccus, rather than "Friedlander's bacillus", resulting in a heated debate between these two researchers. Gram (1884) subsequently demonstrated how pneumonia could result from more than one organism, however it was Weichelselbaum who finally resolved this polemic. His (1986, 1886a, 1886b) culture and pathogenic descriptions of organisms isolated from 129 fatal cases of pneumonia ensured that Fraenkel was attributed with the identification of the pneumococcus as the major cause of lobar pneumonia. On 94 occasions cocci were identified, at the time named *Diplococcus pneumoniae*. In the remaining cases, disease was attributed to a bacillus, named *Bacillus pneumoniae*; today known as *Klebsiella pneumoniae*. It is with some irony that the Gram stain has since become a fundamental test for the differentiation between gram-positive pneumococci and gram-negative *K. pneumoniae*.

The pneumococcus was readily isolated from other bodily sites and rapidly became established as a cause of cerebral meningitis, nephritis, endocarditis, pleuritis, pericarditis, otitis media, as well as causing infections of the accessory sinuses, joints, gums and tonsils and as a cause of otitis media (White, 1938). Its wide association with disease meant that in the 15 years that followed Klebs' (1875) first observations of this pathogen, over 50 researchers had contributed descriptions to its biology, morphology, and pathogenic potential.

1.1.2 Serotherapy

In the early 1890's, the Klemperer brothers demonstrated how rabbits inoculated intravenously with heat-treated pneumococci could acquire a degree of immunity to subsequent pneumococcal infection (Klemperer, 1891b,

Klemperer, 1891a, Klemperer, 1892). Serum obtained from animals that had acquired such immunity was found to possess curative properties against pneumococcal disease when injected intravenously: being successfully used to treat 21 of 23 clinical cases of pneumococcal disease by Pane (1897) and a further 6 pneumococcal patients by Washbourn (1897). Immunity was associated with the formation of "plaquettes" in the serum (Metchnikoff, 1891), which developed over the course of infection in pneumococcal patients, and to which Bezançon and Griffon (1897) ascribed the term "agglutination". Agglutination was not found to be a universal phenomenon. It appeared that distinct pneumococcal races existed within the species, against which immunity was race-specific (Bezançon and Griffon, 1897).

Neufeld and Haendel identified two groups of pneumococci in 1910. In an animal model it was found that sera obtained following infection with "typical" group isolates protected mice against subsequent infection with other "typical" isolates, but not against isolates that they ascribed to an "atypical" group. The reverse was similarly found to be true. These experiments were confirmed by the complementary activities of these groups in agglutination assays (Neufeld and Haendel, 1910). It appeared that effective pneumococcal treatment could be achieved by developing antisera to all pneumococcal types (Neufeld and Haendel, 1910), types that would later become known as serotypes (from serum type). A greater range of methods for group delimitation soon became available, from agglutination, precipitation, complement fixation, and mouse protection. During this period, the lytic effect of bile on the pneumococcus was identified (Neufeld, 1900), which would become an important diagnostic tool for serotype identification in the 1930s (Neufeld and Etinger-Tulcznska, 1930, Austrian, 1999). Sample collections were also becoming larger and more varied. Resultantly, additional pneumococcal types were rapidly being identified: Dochez and Gillespie delimiting 4 groups in 1913 and a further 29 types were described by Cooper and colleagues in 1929.

The method of serotherapy, the intravenous injection of antisera into a patient, had become popular for treating diphtheria during the early 1900s, and

researchers became interested in applying the same technique to pneumococcal disease (Felton, 1924). Dochez and Gillespie (1913) had recognised that pneumococcal types differed in abundance, with 60% of their isolates belonging to two of the four groups that they had identified. The comparative rarity, and greater association with case fatalities of certain pneumococcal types became more evident in the following years (Lister, 1913, Dochez and Avery, 1915). Building on Neufeld and Haendel's observations (1910), the first attempts to treat pneumococcal disease by serotherapy at the Rockefeller Institute specifically targeted the more virulent and prevalent strains of types I and II (Avery et al., 1917). An expansion of serotherapy treatment would later lead to reductions in serotype-specific fatalities by up to 20% (Avery et al., 1917, Austrian, 1999).

Serum sickness, resulting from cross-reactivity between type 14 pneumococcal equine sera and blood group A proved an initial complication for serotherapy (Beeson and Goebel, 1939). Although this was overcome by replacement of equine serum with rabbit serum in the late 1930s, the era of serotherapy had began to draw to a close as antimicrobial treatments began to gain in popularity.

1.1.3 Antimicrobials

Antimicrobials were slow to impact on pneumococcal treatment. Quinine, found to be an effective anti-malarial, showed limited lethality towards the pneumococcus. A related compound, ethylhydrocupreine hydrochloride, colloquially optochin (Morgenroth and Levy, 1911), displayed potent anti-pneumococcal activity in mouse models (Moore, 1915). However, first used to treat pneumococcal disease at the Rockefeller Institute Hospital, its clinical usage was short owing to its high association with blindness (Moore and Chesney, 1917). Furthermore, pneumococci were found to rapidly reduce their susceptibilities to this drug. This was first noted in mouse experiments of 1912, and later in two reports by Moore and colleagues between 1917 and 1918 (Austrian, 1999).

Sulphanilamide, lacking the popularity of serotherapy at the time, was one of the earliest antimicrobials to successfully be used to treat pneumococcal pneumonia (Watson et al., 1993). A more effective derivative, sulphapyridine followed (Whitby, 1938). Evans and Gaisford, Birmingham, England (1938), demonstrated the ability of sulphapyridine treatment to reduce lobar pneumonia case-fatality rates from 27 to 8%, helping sulphapyridine treatment grow in popularity for a short period. However, intolerances to oral administration, erratic absorption rates through the gastrointestinal tract, and a lack of direct relationship between dosage and blood level proved problematic (Taplin et al., 1940). Furthermore, like optochin, sulphapyridine resistance was very quickly reported *in vitro* (MacLeod and Daddi, 1939) and clinically (Ross, 1939, Hamburger et al., 1942, Tillett et al., 1943). Mouse experiments furthermore indicated sulphapyridine resistance was maintained in the population, in the absence of treatment, and despite repeated passage through mice (Frisch et al., 1943, Schmidt et al., 1942).

Although penicillin had been discovered prior to sulphapyridine, the curative potential of this drug was not recognised until the end of the Second World War. It was first described in terms of its ability to improve the isolation of other bacteria, such as *Bacillus influenzae* from mixed samples (Fleming, 1929). The difficulty in scaling up production, and the ready availability of sulphanilamide at the time meant that penicillin was received more as a curiosity than as an antimicrobial suitable for clinical application.

Dubos and colleagues (1939a, 1939b) reignited interest in penicillin following his testing of the antimicrobial properties of gramicidin, a naturally occurring antimicrobial, isolated from spore bearing soil bacilli. Although *in vitro* activity was strong, gramicidin was ultimately found to be unsuitable for clinical use: gramicidin being highly toxic to mice (Hotchkiss and Dubos, 1940) and dogs (MacLeod et al., 1940). This study however inspired Chain and colleagues (1940) to carry out a systematic review of other naturally occurring compounds, shifting the focus from sulphonamide derivatives, which had previously been in vogue. Crucially, this review included penicillin (Chain et al., 1940). Improved culture methods, increased yields, and better assays to test its inhibitory power allowed

for a more thorough investigation into the efficacy of penicillin than had been achieved previously (Abraham et al., 1941).

Abraham and colleagues reported the first clinical uses of penicillin in 1941, where it was used to treat a number of life threatening gram-positive diseases. Penicillin was found to possess "some properties unknown in any antibacterial substance" that had been described prior to this in terms of its efficacy, low toxicity to tissue cells, and bacteriostatic properties (Abraham et al., 1941). The advantages of penicillin over sulphanilamide were clear: having a greater per unit potency, reduced effect of inoculum size on efficacy, and minimal interference of the products of protein hydrolysis (Abraham et al., 1941). Penicillin production was rapidly scaled up in the USA, owing to a lack of pharmaceutical support in the UK, which was financially burdened by the Second World War at the time. Demand for this new highly effective antimicrobial, boosted by the increasing occurrence of sulphonamide resistance (Keefer et al., 1943), began to increasingly outstrip supply. A large study of 500 cases by Keefer (1943) helped establish the efficacy of this drug for treating gram-positive pathogens, while a study by Tillett and colleagues (1944) reported on the use of penicillin to treat 46 cases of pneumococcal pneumonia, helping to better define drug administration. By this time penicillin was rapidly emerging as the drug of choice for the treatment of pneumococcal disease.

The successes that followed clinical introduction of penicillin for the treatment of pneumococcal disease meant that interest in such diseases promptly declined. The growing availability of alternative effective antimicrobials such as cephalosporins (Murdoch et al., 1964, Thornton and Andriole, 1966), tetracyclines (Duggar, 1948), chloramphenicols (Gottlieb et al., 1956), and macrolides (Austrian and Rosenblum, 1953) further acted to reassure, and reduce concern about pneumococcal disease.

1.1.4 Antimicrobial resistance

Variability in pneumococcal penicillin susceptibility had been identified for some time. Firstly by Abraham and colleagues in 1941, where one group of clinical isolates was apparently 30 times less susceptible to penicillin than in another case, both groups carrying a 19F capsule (Abraham et al., 1941, Watson et al., 1993). *In vitro* resistance had also been generated in 1943 (Schmidt and Sesler, 1943, McKee and Houck, 1943) and again in 1945 (Eriksen, 1945). However, clinical resistance took over 20 years to emerge.

By the 1960s, antimicrobial resistance was being increasingly reported clinically for erythromycin, lincomycin (Kislak et al., 1965), tetracycline (Evans and Hansman, 1963), chloramphenicol (Cybulska et al., 1970), and fluoroquinolones (Chen et al., 1999). Intermediate penicillin resistance (0.1-0.2μg/mL) was first reported in two of 200 clinical pneumococcal isolates assessed by Kislak and colleagues, Boston in 1965, although the significance of this observation appears to have been overlooked at the time.

It was Hansman and colleagues (1967) however who reported on the first clinically penicillin resistant pneumococcus, which had a minimum inhibitory concentration (MIC) of 0.6μg/mL to this drug. It was isolated from the sputum of a hypogammaglobulinemia patient in Australia. Within ten years, penicillin resistance had spread globally within the pneumococcal population: reported in the USA (Naraqi et al., 1974), Canada (Dixon et al., 1977), South Africa (Jacobs et al., 1978), England (1977), and Iceland (1989) (Appelbaum, 1987, Appelbaum, 1992). By 1980 approximately 33% of strains identified in New Guinea displayed penicillin resistance (Gratten et al., 1980), and over 10% of isolates were penicillin resistant in Israel, Spain, Poland, South Africa, and Alaska, as well as the cities of New Mexico, Massachusetts, Oklahoma, and Colorado (Appelbaum, 1987). The increasing levels of antimicrobial resistances worldwide meant that there was a renewed interest in prophylactic methods of control.

1.1.5 Whole cell vaccines

The first major investigations into prophylactic methods for preventing pneumococcal disease were driven by the gold mining industry in South Africa. In 1886, vast gold reserves were discovered in the Transvaal region of South Africa, where Johannesburg would later be located. The gold mining industry grew rapidly, with the native African workforce numbering 111,967 in 1899 (Austrian, 1978). Pneumococcal disease became a major cause of illness among the miners, due to the damp and overcrowded working conditions, as well as a system of short-term employment used by the mining industry; this ensured a constant supply of susceptible individuals (Austrian, 1978). The overall burden of disease becoming so great that authorities in Britain threatened to close the mines at the time (Austrian, 1999). In an attempt by the mining companies to alleviate the situation, the assistance of Sir Almroth Wright was sought- highly praised for his then recent work on anti-typhoid fever inoculations. Travelling to South Africa in 1911, Wright conducted a number of vaccine trials there in an attempt to control pneumococcal disease. This whole-cell vaccination programme would eventually include over 50,000 miners (Wright et al., Austrian, 1999), and was declared a success in Wright's own assessment: vaccination appeared to reduce pneumococcal disease in the first 2 months of administration, although case fatality rates were not affected (Grabenstein and Klugman, 2012). However, with Wright being a famous opponent of statistical analyses, it is unclear whether vaccination was truly effective (Maynard, 1913, Austrian, 1999). Although a number of whole cell vaccines were available at this time, little was known about serotypes. As such vaccines were not developed to target the most prevalent serotypes making it unlikely that any strong clinical reductions in pneumococcal disease would have been observed following vaccination at the time.

On Wright's departure, Lister continued the vaccination programme in South Africa. With the increasing understanding of the importance of serotypes, Lister (1916) delimited and developed whole cell vaccines against the serotypes predominantly causing disease. The forward–thinking Lister (1916) also rationalised that vaccination of a proportion of the population could provide a

protective effect to the rest of the population, which would later become known as herd immunity. However, his experimental trials were poorly constructed, and failed to convince people of the efficacy of vaccination at the time (Heffron, 1939, Austrian, 1999, Grabenstein and Klugman, 2012). Influenza disease increased rapidly during the First World War, peaking with the influenza pandemic of 1918. An estimated 20 to 100 million people are thought to have died worldwide during this epidemic, secondary infection with pneumococci making a significant contribution to this mortality (Chien et al., 2012). As a result, a wide variety of "influenza" vaccines became available. The aetiological agent of influenza was however unknown at this time, and the high association between pneumococcal disease and influenza meant that vaccine developers often erroneously attributed influenza disease to the pneumococcus. Consequently pneumococcal components were often included in the vaccines of this period (Austrian, 1978, Grabenstein and Klugman, 2012).

Dedicated pneumococcal vaccines were also available during this epidemic. Chien and colleagues (2010) recently re-analysed data from civilian and military studies into mixed killed pneumococcal vaccines used between 1918 and 1919. Such vaccines, although highly variable in potency, did appear to be effective, with pneumonia cases reduced by between 19 and 70%, and reducing case fatality rates among influenza patients by between 18 and 82% (Chien et al., 2010).

1.1.6 Polysaccharide vaccines

The identification of the capsule as an important target for the host's adaptive immune response aided further vaccine development (Dochez and Avery, 1917, Heidelberger and Avery, 1923, Schiemann and Casper, 1927, Francis and Tillett, 1930, Dubos and Avery, 1931). Vaccine design began to shift away from whole cell vaccines but towards purified capsular polysaccharide types. Felton carried out a number of largely successful trials of polysaccharide vaccines during the 1930s helping to establish the efficacy of serotype specific polysaccharide vaccinations (Felton et al., 1941, Grabenstein and Klugman, 2012). Macleod and

colleagues conducted a large-scale trial of a polysaccharide vaccine in 1945. Here, vaccination conferred a level of type specific immunity (against types 1, 2, 5, 7), which they suggested lasted for a minimum of 6 months. In addition a reduction in type-specific carriage and pneumococcal pneumonia was found in the non-immunised control group, confirming Listers (1913) earlier predictions (Austrian, 1999, MacLeod et al., 1945). Additional trials by Kaufman (1947), demonstrated a 90% reduction in pneumococcal type specific (types 1, 2, and 3) pneumonia and bacteraemia. This culminated in the granting of a licence to E. R. Squibb and Sons for two hexavalent polysaccharide pneumococcal vaccines: one for adults, protecting against serotypes 1, 2, 3, 5, 7 and 8, and one for children, protecting against serotypes 1, 4, 6, 14, 18 and 19 (Grabenstein and Klugman, 2012). Introduced commercially in 1947, uptake was limited during this period, owing to the availability of the then highly effective and popular antibiotics, which included penicillin, chlortetracycline, and chloramphenicols. As a consequence the vaccine was withdrawn from sale in 1951, and the license voluntarily revoked in 1954 (Grabenstein and Klugman, 2012).

Vaccines began to be revisited largely due to the work of Austrian in the 1960s, coupled with the emergence of antibiotic resistance. Austrian and Gold's 1964 publication on pneumococcal bacteraemia identified the high incidence of mortality and clinical sequelae that developed in patients despite the use of antimicrobials. They argued that antimicrobials needed to be supplemented with prophylactic and therapeutic measures, and called for a return to serotyping of pneumococci to aid epidemiological studies (Austrian and Gold, 1964).

By this time, more pneumococcal serotypes had been recognised than could conceivably included in a vaccine. However, enhanced surveillance data from Europe, North America, and South Africa revealed that approximately 18 serotypes accounted for seven-eights of bacteraemic illnesses (Austrian, 1999). Vaccine development was therefore focussed on including as many of these prominent serotypes as possible.

Clinical trials of hexavalent and 12-valent pneumococcal vaccines were coordinated by Austrian (1976), in ~12,000 gold miners of the Transvaal region of South Africa, leading to an 80% reduction in vaccine type pneumonia. Further clinical trials lead by Pieter Smit in the 1970s demonstrated safety, and a 92% reduction in pneumococcal pneumonia using the 12 valent vaccine, and a 76% reduction in pneumococcal pneumonia using a hexavalent vaccine (Grabenstein and Klugman, 2012). Similarly successful vaccine efficacy trials were conducted in Papua New Guinea by Riley and colleagues (1977, 1991).

Greater evidence for the safety and efficacy of vaccines, coupled with reduced antibiotic susceptibility lead to a more permissible environment for vaccination than had occurred previously. As such, in 1977 a tetradecavalent (14-valent) pneumococcal polysaccharide vaccine (PPV) was licensed for use in South Africa. Later this was expanded to a 23-valent PPV, which was introduced in 1983, and targeted 87% of bacteraemic pneumococcal disease serotypes then circulating in the US (Austrian, 1999).

In the 1970s however it was recognised that there was an inadequate antibody response to pneumococcal polysaccharides in children below 2 years of age, yet this represented a significant risk group for pneumococcal disease (Grabenstein and Klugman, 2012).

1.1.7 Conjugate vaccines

Avery and Goebel had generated the first conjugate pneumococcal vaccine in 1929. Here it was found that immunogenicity to type 3 polysaccharide was enhanced in rabbits following the attachment of this polysaccharide to horse serum globulin (Avery and Goebel, 1929). However, it would not be until much later that the greater potency of conjugate vaccines for generating adaptive immunogenicity in children was recognised.

Research into *Haemophilus influenzae* conjugate vaccines preceded the development of a similar pneumococcal type. Attachment of *H. influenzae*

type B (HiB) polysaccharide to the diphtheria toxoid indicated enhanced antibody responses on subsequent antigen exposure in children (≤6 months)(Ward et al., 1988). Immunogenicity to various carrier molecules was also assessed, with significant differences in the levels of immunogenicity generated (Granoff et al., 1992). Following the licensing of a HiB conjugate vaccine in 1989, an approximate 98% reduction in this disease was observed in the USA (CDC, 1996b, Adams et al., 1993, Barbour et al., 1995).

Such successes in *H. influenzae* prophylaxis fuelled interest in the development of a similar vaccine for pneumococci, which, owing to the reductions in *H. influenzae* disease, had become the leading cause of meningitis among children in the USA and to which antibiotic resistance was increasing worldwide. Development however was complicated, HiB being based on the single type B polysaccharide, compared to the over 90 pneumococcal serotypes that had by now been recognised.

Robbins and Schneerson carried out a number of studies into the efficacy of pneumococcal conjugate vaccines in the early 1980's, seeming to favour a tetanus conjugate (Schneerson et al., 1984, Robbins and Schneerson, 1990). Enhanced immunity of conjugate pneumococcal vaccines was similarly demonstrated among children compared to purified polysaccharide versions (Steinhoff et al., 1994, Kayhty et al., 1995). Further investigations into the variations in immunogenicity generated using difference carrier molecules favoured the use of an inactivated diphtheria toxin carrier, CRM_{197} (Anttila et al., 1999). A pentavalent conjugate of this type was found to be safe and effective in children when delivered at 2, 4 and 6 months of age (Ahman et al., 1996, Daum et al., 1997). Subsequent expansion to include seven of the most prevalent pneumococcal serotypes in disease, lead to the development of the pneumococcal conjugate vaccine (PCV) 7. This vaccine was shown to be similarly effective, and offered the potential to reduce invasive pneumococcal disease among US children by 85% (Rennels et al., 1998).

1.1.7 Vaccine Scale-Up

By 1996, the Centre for Disease Control and Prevention (CDC) estimated *S. pneumoniae* resulted in 7,000,000 cases of otitis media, 500,000 cases of pneumonia, 50,000 cases of bacteraemia, and 3,000 cases of meningitis in the USA, leading to a drive for vaccine introduction. Concerns over the effectiveness of vaccination were however also raised. Unlike polysaccharide vaccines, conjugate vaccines were found to reduce carriage of vaccine type serotypes in the nasopharynx (Spratt and Greenwood, 2000). The serotype-specificity of these vaccines therefore lead to concern that vaccine-type serotypes would be replaced by other serotypes present in the pneumococcal population (Lipsitch, 1999, Spratt and Greenwood, 2000). Serotype switching was also known to occur at the time, suggesting virulent pneumococcal lineages could circumvent vaccine control, by switching to a non-targeted serotype (Spratt and Greenwood, 2000). However, the burden of disease meant mass vaccination was still highly desirable, and that the rates of capsule switching were low enough that vaccination would still be effective at removing the more virulent serotypes, and those most often associated with multidrug resistance from the population (Spratt and Greenwood, 2000).

PCV7 was licensed for use in the USA and Europe in 2000, being rolled out across additional countries the following year. Where introduced, this vaccine was routinely administered to children and infants during the first 2 years of life (Paradiso, 2011).

1.1.8 Vaccine windfall

Routine vaccination of children lead to a rapid and sustained reduction in vaccine type invasive pneumococcal disease (IPD) cases among children, and unvaccinated adults. Pilishvili and colleagues (2010) reported a 45% reduction in IPD cases across all ages, and a 94% reduction in vaccine type IPD in the USA between 2000 and 2007. Similar reductions in IPD, of 34%, were recorded in England and Wales by 2010 (Miller et al., 2011). A 98% reduction in vaccine type disease occurred across all ages (Miller et al., 2011). In addition herd immunity

was also recorded in unvaccinated infants (Poehling et al., 2006), adults (Whitney et al., 2003), and adults with Human Immunodeficiency Virus (HIV) (Cohen et al., 2010). PCV7 also lead to a reduction in antibiotic usage (Dagan and Klugman, 2008), and reductions in the isolation of penicillin non-susceptible pneumococci from cases of IPD among children (Kaplan et al., 2004) and adults (Kyaw et al., 2006).

However, the prevalence of non-vaccine types, and their occurrences in disease has increased over this period, as was predicted (Steenhoff et al., 2006, Hicks et al., 2007, Pai et al., 2005, Hanage et al., 2011). Serotype switching of virulent vaccine type strains to non-vaccine types, such as 19A, has similarly been identified widely (Moore et al., 2008, Hanage et al., 2011, Croucher et al., 2011).

Although vaccination has proven to be an effective preventative measure against pneumococcal disease (Kyaw et al., 2006), limited serotype coverage, capsule switching events, and strain cycling prevent the current vaccines from being an effective long-term strategy for control of pneumococcal disease. PCVs continue to be developed, with the aim to include further serotypes- with PCV13 recently being rolled out (Paradiso, 2011). The long-term strategy for control of pneumococcal disease remains tenuous, with continued epidemiological surveillance deemed a priority (Spratt and Greenwood, 2000).

1.2 Current populations at risk

1.2.1 Risk factors for disease

Pneumococcal disease occurs predominantly in the young (≤5yrs), elderly (65yrs≤) and immunocompromised (Pastor et al., 1998, Simberkoff et al., 1984, Lynch and Zhanel, 2010). For children, immune naivety, reducing the effectiveness of clearance, and enhanced contact with other colonised individuals, such as occurs in nurseries, promotes pneumococcal carriage, a prerequisite for invasive disease (Gray et al., 1982, Bogaert et al., 2004, Usuf et al., 2014). Likewise, pneumococcal clearance is reduced among the elderly due to an age associated decline in immune-effectiveness, termed immunosenescence.

This has more recently been attributed to a change in the mucosal immune regulation (Krone et al., 2013).

It is less clear whether colonisation rates are higher among HIV infected individuals, or whether such individuals are more susceptible to colonisation by the more virulent strains (McNally et al., 2006, Janoff et al., 1993, Polack et al., 2000, Madhi et al., 2007, Nunes et al., 2013).

Increased IPD risk is similarly associated with a variety of congenital and chronic medical conditions (Muller et al., 2005, Bogaert et al., 2004, Barrett-Connor, 1971, Powars et al., 1981), ethnicity (Trotman et al., 1995, Watson et al., 2006, Chen et al., 1998, Pastor et al., 1998, Robinson et al., 2001), and a number of socioeconomic factors, such as income, day-care attendance, household crowding, and the occurrence of smoking in the household (Chen et al., 1998, Powars et al., 1981, Nuorti et al., 2000, Cardozo et al., 2008, Bakhshaee et al., 2012).

Untreated pneumococcal pneumonia has been estimated to result in an adult case fatality rate of 30-35% (Heffron, 1939), whereas case fatality rates for pneumococcal meningitis in children have been estimated to be in the region of 20% and 50% in developed and undeveloped countries respectively (Koedel et al., 2002). The level of morbidity associated with pneumococcal disease is also significant. Clinical sequelae occur in up to 10% and 30% of children in developed and developing countries (Koedel et al., 2002), manifest in terms of neurophysical, or neurological impairment, and hearing loss.

As such, the annual cost of pneumococcal disease remains high, despite preventative measures. An estimated €10.1 billion are spent annually on pneumococcal pneumonia in Europe, with indirect costs, such as lost work, accruing to more than €3.6 billion (Fedson et al., 2011). The direct costs of IPD in the USA are estimated at US$3.7 billion each year, and an additional US$1.8 billion amassing from indirect costs (Fedson et al., 2011). Community Acquired Pneumonia (CAP) is similarly estimated to cost over US$17 billion annually

(Fedson et al., 2011). However, the burden of pneumococcal disease is predominantly felt in developing countries.

1.2.2 The Meningitis Belt

The incidence of pneumococcal disease occurs predominantly in developing countries (Gavi, 2013)(Figure 1). This is largely due to widespread malnutrition and associated deterioration in immune function (Rodríguez et al., 2011), and a high incidence of diseases that facilitate pneumococcal infection, particularly HIV and sickle cell disease. HIV depletes CD4+ T cells contributing towards B cell dysfunction, and reduced antibody response- IgA being particularly important for the clearance of pneumococci from mucosal sites (Hart et al., 2007, Moja et al., 1997, Zhang et al., 2015). Sickle cell anaemia in contrast leads to the chronic up-regulation of the Platelet Activating Factor (PAF) receptor, PAFr, in response to inflammation. Through the action of Choline binding protein (ChoP) pneumococci are able to bind PAFr facilitating transcytosis and invasion (Miller et al., 2007).

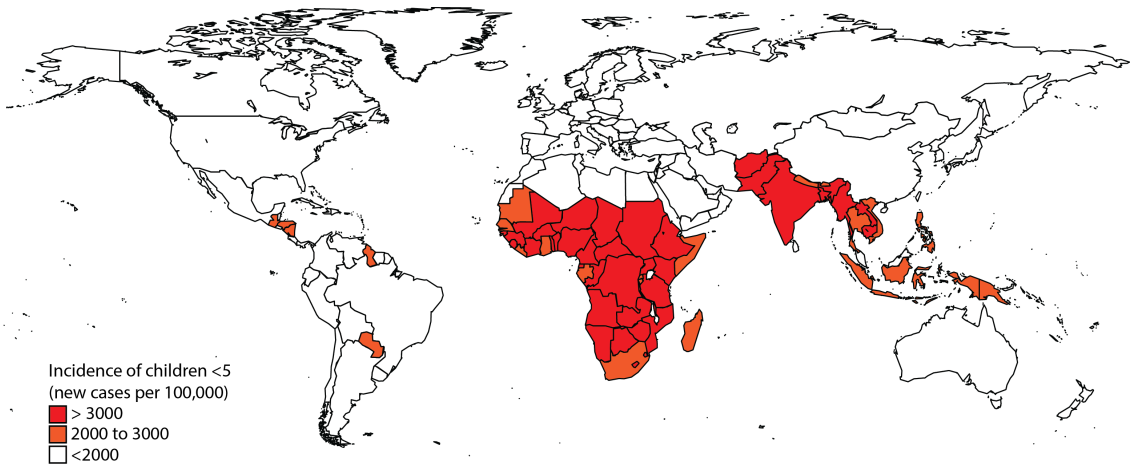


Figure 1: Incidence of childhood pneumonia globally (Gavi, 2015).

The "Meningitis belt" (Lapeyssonnie, 1963) stretches from Senegal to Ethiopia (Figure 2), and demarks a region of high endemic levels of meningitis, caused predominantly by *Neisseria meningitidis* and the pneumococcus. Large epidemics of meningitis occur periodically, but unpredictably in this region during the dry season every 2 to 10 years (Trotter and Greenwood, 2007). Waves of severe

meningitis epidemics have also occurred outside of the Meningitis Belt since the late 1980s (Molesworth et al., 2002). As such, Moleswoth and colleagues (2002) redefined African areas of risk outside of the belt, to include Sahel, the East African Rift Valley, the Great Lakes region and part of South Africa extending between Mozambique to Angola and Namibia (Molesworth et al., 2002)(Figure 2). Enhanced surveillance, and the use of epidemic thresholds for early detection were proposed by the WHO in an effort to improve disease control in this area (WHO, 2000, Lewis et al., 2001, Molesworth et al., 2002). Pneumococcal outbreaks are in part a consequence of the unusually virulent serotype 1 serotype, which although rarely carried, is one of the leading causes of IPD in many areas of the world. Interestingly a striking geographic structure has been found to occur in this serotype, one lineage being present in Europe, the United States, and Canada, another being found in Africa and Israel, and a third lineage apparently restricted to South America (Brueggemann and Spratt, 2003). The importance of serotype 1 in IPD lead to its inclusion in the newer conjugate vaccines.

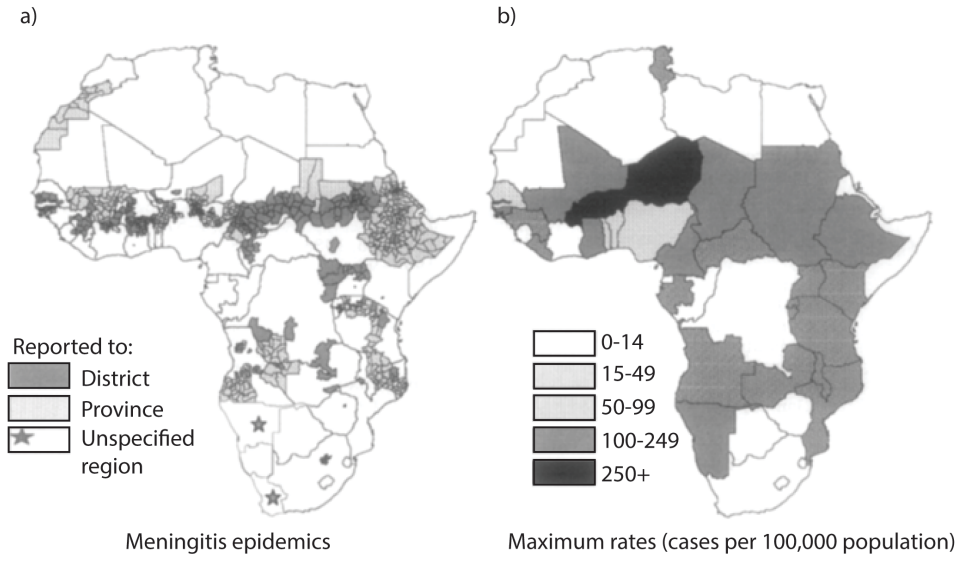


Figure 2: The Meningitis Belt delimits a region in which meningitis occurs in periodic epidemics. Molesworth and colleagues redefined the extent of the Meningitis Belt based on case detection (a) and incidence of disease (b) in 2002 (Molesworth et al., 2002).

The recent establishment of the Global Alliance for Vaccines and Immunisation (GAVI) aims to challenge vaccine preventable disease in previously largely overlooked communities. This alliance sponsors the Pneumococcal vaccines

Accelerated Development and Introduction Plan (PneumoADIP), which aims to expedite access to pneumococcal vaccines to children globally (Gessner et al., 2010). An estimated 1-4 million episodes of pneumococcal pneumonia occur annually in Africa, with ~800,000 pneumococcal child (<5yrs) deaths occurring within developing countries each year (Scott, 2007). Facilitated by the establishment of GAVI and PneumoADIP, there has been a large-scale rollout of PCVs within the developing world in recent years.

1.2.3 Malawi

Malawi, a GAVI-eligible country, located in southern Africa is an impoverished country that has suffered heavily from widespread Acquired Immunodeficiency Syndrome (AIDS) epidemics since 1975 (Gordon et al., 2000). HIV sero-prevalence in antenatal women has increased 16-fold to date (Taha et al., 1998), which has been further associated with a 4-fold increase in medical admissions to Queen Elizabeth Central Hospital (QECH)- Malawi's largest hospital, located in Blantyre (Gordon et al., 2000). Out of a population of over 14 million, an estimated 900,000 children and adults live with HIV, resulting in over 85,000 deaths per annum (Makombe et al., 2006). The high incidence of immunosuppressive disease has lead to an increase in morbidity and mortality caused through secondary infection with pneumococcal bacteria (Gordon et al., 2000).

Prior to the introduction of free Anti-Retroviral Therapy (ART) pneumococcal disease was responsible for 65% of mortality associated with meningitis, and 20% for pneumococcal pneumonia at QECH (Gordon et al., 2002). IPD and lobar pneumonia are similarly highly correlated with HIV, having a 90% co-infection rate (Jones et al., 1998, Gordon et al., 2000, Carrol et al., 2007). Patients with HIV are further estimated to suffer a 30-100 times greater risk of IPD compared to age-matched healthy controls (Janoff et al., 1992, Jones et al., 1998). ART was introduced in 2004 in an attempt to tackle HIV induced mortality and the

incidence of pneumococcal disease has since been reduced substantially (Everett et al., 2011).

It is hoped that the introduction of PCV13 in 2011 will lead to sustained declines in IPD. Reassuringly, PCV13 has been predicted to have a protective effect against 62.9% of all current circulating strains, in a population where serotype 1 is highly prevalent in disease, at 20.5% (Table 1)(Everett et al., 2012).

It remains too early to tell how effective PCV13 introduction will be however (Wall et al., 2014), and there are already indications of the ability for serotype switching to occur naturally in Malawi, an effect which could be exacerbated by vaccine pressure (Kamng'ona et al., 2015, Everett et al., 2012). Case fatality rates remain high, even among those with access to timely and effective healthcare, and invasive pneumococcal disease still accounts for between 8 and 50% of meningitis fatalities in Africa (French et al., 2010). Therefore, the need for continued serotype surveillance, and the availability of effective antimicrobial regimes remains important, helping to achieve the reductions in child mortality that were outlined in the Millennium development goals (Nyirenda et al., 2014).

Penicillin resistance remains low in Malawi, rising only marginally between 2000 and 2009 from 14.7% to 15.9% respectively, based on meningitis breakpoints (Everett et al., 2011). During this time penicillin has been used as a first-line defence against pneumonia and sepsis (Cornick et al., 2011). In addition ceftriaxone has been used as a first-line treatment against meningitis since 2004 (Cornick et al., 2011). The recent occurrence of ceftriaxone resistance (pers. comm. Everett, 2015), following a sustained creep in loss of susceptibility (Everett et al., 2011) further raises concerns as to the long-term efficacy of current treatments. This threat means that there is a growing necessity to identify alternative therapies that can be invoked once resistance breakpoints are reached.

Vaccine	Age (years)	Serotype coverage
PPV23	>2	1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19A, 19F, 20, 22F, 23F, 33F
PCV7	<5	4, 6B, 9V, 14, 18C, 19F, 23F
PCV13	<5	1, 3, 4, 5, 6A, 6B, 7F, 9V, 14, 18C, 19A, 19F, 23F
Serotypes currently circulating in Malawi		1, 6A/6B, 14, 23F, 12F, 19F, Sg18, 4, 7F, 10A, 16F, 33F, 35F, 3, 5, 9V, 14/4, 15a, 19a, 7c, 8, 34, 11a, 17f, 35B

Table 1: Serotype coverage offered by the currently available pneumococcal vaccines. PPV23 currently offers the broadest protection against pneumococcal serotypes, but being a polysaccharide based vaccine it has a low efficacy in children. The conjugate vaccine PCV7 was introduced widely in the developed world when mass vaccination was first undertaken around the turn of the millennium (section 1.1.8). The wider serotype coverage offered by the newer PCV13 is predicted to protect against of 60% of circulating serotypes in Malawi. The bottom row indicates those serotypes currently prevalent in Malawi for comparison with those covered in PCV13 (modified from Cornick et al., 2011).

1.3 Pneumococcal biology

1.3.1 Identification and Morphology

Few species have undergone as many name changes as *S. pneumoniae* (Table 2), owing to its association with multiple infections and diseases, and differing morphological characters. Identification remains problematic, with a variety of methods employed diagnostically, primarily to differentiate pneumococci from other co-colonising alpha-haemolytic bacteria present in mixed samples. Colonies are generally grown on agar plates supplemented with blood as a source of catalase, where they form alpha-haemolytic colonies. Alpha haemolysis results from the reduction of haemoglobin to methemoglobin among the red blood cells surrounding pneumococcal colonies (Buxton, 2005), leading to a green halo effect. The drug optochin, which never achieved success as a clinical antimicrobial, is now used as part of the standard pneumococcal diagnostic test: optochin susceptibility being used for the presumptive identification of pneumococci. A 5µg/mL optochin disc should yield a 14mm zone of inhibition among pneumococcal colonies (CDC, 2012). Bile (sodium deoxycholate) similarly lyses pneumococcal

cells, allowing for their differentiation from other bile-resistant alpha haemolytic bacteria (CDC, 2012).

At the microscopic level, the pneumococcus tends to form diploid lancet shaped colony clusters, rather than the chains of adjoined cocci formed by other streptococci. Individual bacteria are gram-positive and may or may not possess an outer polysaccharide capsule.

Name	Year adopted	Reference
Monas pulmonale	1875	Klebs, 1875
Microbe septicemique du saliva	1881	Pasteur, 1881
Pneumoniekokken	1883	Matray, 1883
Coccus lanceole de la pneumonie	1883	Talamon, 1883
Pneumoniemikrokokken	1883	Friedlander, 1883
Pneumonie-Micrococcen	1883	Friedlander, 1883
Micrococcus Pasteuri	1885	Sternberg, 1885
Pneumoniemikrococcus	1886	Fraenkel, 1886
Pneumococcus	1886	Fraenkel, 1886
Bacillus septicus sputigenus	1886	Flugge, 1886
Diplokokkus lanceolatus pneumoniae	1886	Flugge, 1886
Diplococcus pneumoniae	1886	Weichselbaum, 1886
Bacillus salivarius septicus	1887	Biondi, 1887
Micrococcus pneumoniae crouposae	1887	Sternberg, 1887
Streptococcus lanceolatus Pasteuri	1888	Gamaleia, 1888
Diplococcus lanceolatus	1888	Foa, Bordoni-Uffreduzzi, 1888
Virus pneumonico	1889	Gabbi, 1889
Diplococcus pneumoniae	1886-proposed, popular in 1930s	Weichselbaum, 1886
Streptococcus pneumoniae	1972	Wannamaker and Matsen, 1972; Buchanan and Gibbons, 1974

Table 2: A list of the various names that have been assigned to the pneumococcus since its initial discovery. Also indicated is the year the name was proposed, and the proposing author(s) (White, 1938).

1.3.2 The Pneumococcal Genome

The first whole genome pneumococcal sequence was published in 2001, which indicated a genome size of approximately 2.2mbp, and GC content of 40%- both figures remaining relatively consistent with later genomic studies, the smallest genomes being approximately 2mbp (Keller et al., 2013). The genome was found to contain an estimated 2,200 coding regions, of which 64% have been assigned to a biological role (Tettelin et al., 2001).

The introduction of high-throughput sequencing technologies in 2005 allowed for more complex analyses to be performed through comparative genomics, availing of the increasing numbers of whole genome sequences

available for study (Figure 3). This gave rise to the concept of a pan and core genome to describe genetic variation within a species: core genes indicating those that are conserved across the species, whereas the accessory component contains dispensable genes (Tettelin et al., 2005, Donati et al., 2010). The most recent estimates suggest that 1,647 of the 2,200 genes present are contained within the core genome, with the rest assigned as accessory (Donati et al., 2010). Transcriptomics, based on sequencing the mRNA is likely to lead to further insight into the role of such genes under different environmental conditions (Song et al., 2009, Ogunniyi et al., 2012).

Recombination has been found to play a predominant role in pneumococcal genome rearrangement and evolution (Donati et al., 2010). The relative likelihood of a nucleotide polymorphism having been introduced through recombination rather than mutation has been estimated at between 7.2 (Croucher et al., 2011) and 66 (Feil et al., 2000), based on whole genome (single lineage) and housekeeping genes (species wide) respectively. Disruption of recombination machinery has similarly been found to correlate with a rapid decline in previously successful pneumococcal lineages (Croucher et al., 2014b). Inter-species recombination with other closely related bacteria present during nasopharyngeal colonisation has been found to play an important role in genome evolution of the pneumococcus.

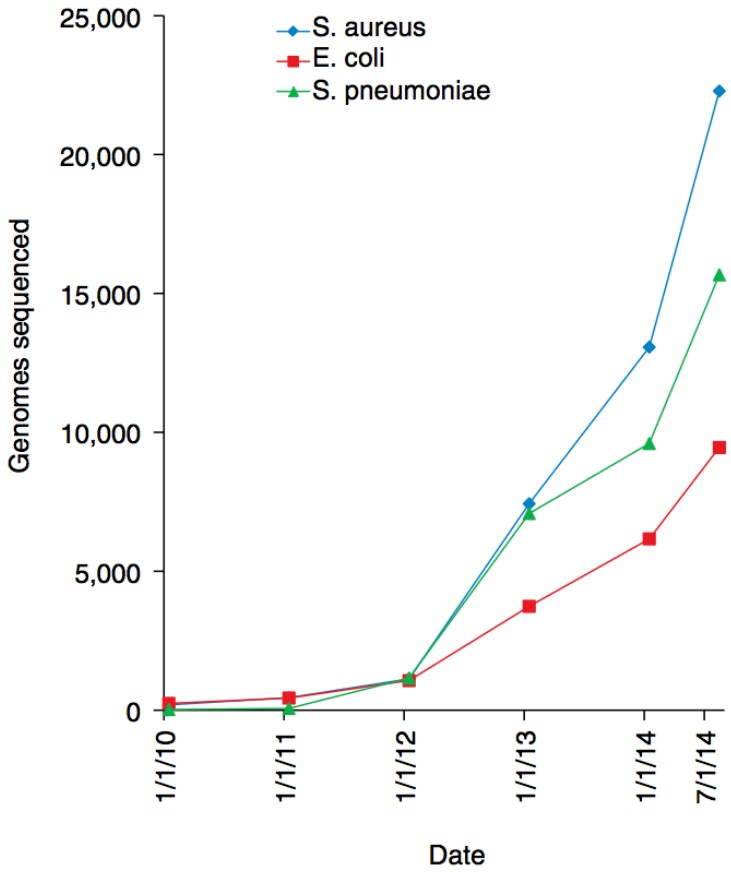


Figure 3: Numbers of whole genome sequences available for some of the most commonly studied bacterial species (from Read and Massey, 2014).

1.3.3 Host colonisation

The nasopharynx forms the primary niche of the pneumococcus. The nasopharynx is colonised soon after birth by a diverse, and dynamic microbial community (Faden et al., 1997, Simell et al., 2012). Nasopharnygeal colonisation by pneumococci typically peaks between the ages of 2 and 3 (Bogaert et al., 2004, Henriques-Normark and Tuomanen, 2013), where it occurs in approximately 60% of children (<5yrs) in developed countries (Bogaert et al., 2004, Hussain et al., 2005), but can reach levels of 90% among children in resource-poor countries (Hill et al., 2008). Incidence typically declines thereafter (Usuf et al., 2014), likely reflecting the acquisition of adaptive immunity (Lipsitch et al., 2005).

1.3.4 Capsule

The capsule was the earliest virulence target (Table 3) described for the pneumococcus (Heidelberger and Avery, 1923), with over 94 different variants described, aiding immune avoidance. When present, it forms a polysaccharide coating above the cell wall (Hammerschmidt et al., 2005). Its expression is however tightly regulated, and appears to be suppressed under oxygenated conditions (Weiser et al., 2001, Carvalho et al., 2013).

The capsule helps protect against mucin defences during the early stages of invasion, and by concealing underlying cell wall associated antigens, such as teichoic acids from immune exposure. The capsule carries a negative charge, and it has been suggested that a physiochemical effector mechanism resulting from this charge aids in reducing mucus mediated clearance of pneumococci during the early stages of colonisation (Simell et al., 2012). During the later stages of invasion the binding of positive charged host molecules can help to neutralise this capsular charge (Simell et al., 2012). Expression of a number of surface attached exoglycosidases, namely NanA, BgaA and StrH allow deglycosylation of host molecules. These enzymes work sequentially to expose mannose receptors present on human glycoproteins, by digesting sialic acid, galactose and N-acetylglucosamine (King et al., 2006). Such enzymes help expose novel adherence receptors, inhibit clearance and can provide a source of carbon for growth (Simell et al., 2012).

Capsular expression is generally thought to inhibit adherence to the epithelium, and consequently colonisation efficiency corresponds with a down-regulation in capsule expression. This action exposes the underlying cell-wall receptors and virulence proteins.

The cell wall is largely composed of peptidoglycan interspersed with teichoic acid attached to N-acetylmuramic acid. At its base, lipoteichoic acid attaches to the cell membrane via a lipid moiety. A number of choline binding proteins (CBPs) are present on the surface of the wall, anchored

in place through non-covalent bonds to teichoic acid phosphorylcholine, and membrane bound lipoteichoic acids (McDaniel et al., 1991). Exposure of these CBPs changes the electrostatic and hydrophobic properties of the pneumococci's surface thought to facilitate adherence to host cells (Swiatlo et al., 2002).

Following the action of exoglycosiades, newly exposed N-acetyl-glycosamine carbohydrate residues on the surface of resting epithelial cells act as anchoring points for cell wall associated surface proteins, such as pneumococcal surface adhesion A (PsaA)(Bogaert et al., 2004, Rajam et al., 2008). This establishment of a strain within the host, following the adherence of pneumococci to the epithelial cells lining the respiratory tract, is defined as acquisition (Simell et al., 2012). The duration of carriage after this point depends largely on the ability for the immune system to clear infection, and the actions of other co-colonisers. As such, the duration of carriage is highly variable, lasting from a few days, to up to 40 weeks (Sleeman et al., 2006, Turner et al., 2012).

Co-colonisers of the nasopharyngeal niche have been found to affect the establishment, and maintenance of pneumococcal carriage. In particular other members of the alpha-haemolytic streptococci have been found to inhibit pneumococcal growth in addition to that of other common nasopharyngeal colonisers, such as *H. influenzae, S. aureus,* and *M. catarrhalis* (Holm and Grahn, 1983, Fujimori et al., 1996, Brook and Gober, 2005, Tano et al., 2000). A negative association has also been found to occur between *S. pneumoniae* and *S. aureus*, *H. influenza*, and *M. catarrhalis* colonisation in healthy children and *in vitro* (Bogaert et al., 2004, Regev-Yochay et al., 2004, Madhi et al., 2007). This effect appears to result from the production of hydrogen peroxide by *S. pneumoniae.* The pneumococcus, like other lactobacillales, lacks a complete tricarboxylic acid (TCA) cycle. Glucose is converted to pyruvate through conventional glycolysis, however additional energy must be derived independent of the TCA cycle. Instead, pyruvate appears to be subject to the competing activities of lactate dehydrogenase, pyruvate oxidase (SpxB), pyruvate formate-lyase and a putative

pyruvate dehydrogenase complex (Carvalho et al., 2013). Under aerobic conditions SpxB drives the conversion of pyruvate to acetyl-phosphate and acetate. Whilst this process yields an additional molecule of Adenosine Tri Phosphate (ATP) it results in the production of hydrogen peroxide (Pericone et al., 2000, Taniai et al., 2008). Lacking the normal peroxide scavengers, such as catalase, carried by many other aerobic bacteria, hydrogen peroxide accumulates in the supernatant. Millimolar concentrations of hydrogen peroxide can accumulate as a consequence. Whilst the pneumococcus is surprisingly tolerant to high concentrations of hydrogen peroxide, it is highly toxic to competing bacteria, and to the host tissue, owing to the occurrence of hydroxyl radicals (Pericone et al., 2000).

Aid nasopharyngeal colonisation	Function
Capsule	Inhibits opsonisation, and prevents entrapment within the nasal mucus
ChoP	Binds to PAFr present on the epithelial surface
CbpA (or PspC), SpsA	Binds to secretory component of the human secretory IgA
NanA, BgaA, StrH	Sequential activities of these enzymes cleave terminal sugars from human glycoconjugates, possibly helping to uncovering receptors for adherence
Hyl	Digestion of hyaluronan-containing extracellular matrix components
PavA	Binds the glycoprotein fibronectin present in the extracellular matrix
Eno	Binds to plasminogen (plasmin zymogen)
Competitor inhibitors	**Function**
Bacteriocin (pneumocin)	A proteinaceous toxinthat targets member of the same species
Immune subjugation and infection	**Function**
Ply	A cytolytic toxin capable of competence activation. Important for colonisation and infection in mouse models
PspA	Prevents the binding of C3 complement component to the pneumococcal surface. Also binds lactoferrin- a mutifuncitonal transferrin family protein, found in saliva, tears and milk
LytA	Digests the cell wall, resulting in the release of Ply
PsaA	A divalent metal-ion-binding lipoprotein component of an ATP binding cassette (ABC) transport system, with a specificity for manganese. Involved in resistance to oxidative stress
PiaA, PiuA	ABC transport system component
NanA, NanB	Aids colonisation through digestion of the mucus layer, uncovering receptors for adherence, modifying the surfaces of co-colonising bacteria, and/or performing functional modifications to hosts' clearance glycoproteins
IgA	Cleaves human IgA

Table 3: Commonly identified virulence proteins found among pneumococci, used to aid colonisation and immune subjugation (reproduced from Kadioglu et al., 2008).

Much of what is currently known about host immune responses to pneumococcal infections has been inferred from murin models. Such models suggest an important role for $CD4^+$ T cell-mediated cellular responses in both naïve and secondary infections. Monocyte/macrophage recruitment to the upper respiratory tract mucosal surfaces, facilitated through TLR2, $CD4^+$ T cells, and IL-17A plays a major role in clearance. In contrast, while neutrophils often proliferate and converge at the site of inflammation prior to

monocytes/macrophages, such cells appear to play a minor role in clearance in the naïve host. In contrast, among previously exposed mice, which have acquired a $CD4^+$ T cell memory, an earlier and more pronounced influx of phagocytic cells, particularly neutrophils, expedites clearance. This is perhaps due to the opsonising affect of antibodies in immune mice, which aids phagocytosis by up to fourfold relative to the unexposed state (Ferreira et al., 2013). In a naïve infection, the neutrophilic burst occurs prior to the generation of effective antibodies (Zhang et al., 2009, Verschoor et al., 2014). The effect therefore appears limited, and reliant on complement activation, which predominantly occurs via the classical pathway (Paterson and Mitchell, 2006). Subsequent $CD4^+$ T cell activation, and cytokine secretion promotes neutrophil and macrophage recruitment. Interleukin (IL) 17, secreted by T-helper 17 (Th17) immune cells is particularly important for immune clearance, enhancing monocyte/macrophage migration to the sites of inflammation (Verschoor et al., 2014).

1.3.5 Invasion

Colonisation of the nasopharynx is usually asymptomatic, and rarely results in disease. Whilst resting epithelial cells cannot be invaded, intracellular compartmentalisation can be viewed following inflammatory activation of these cells (Geelen et al., 1993). Inflammation consequently appears to be a prerequisite to invasion and usually occurs in response to the presence of pneumococcal virulence components.

A number of pneumococcal virulence components may be responsible for invasion, such as capsule, LytA, and pneumolysin, and hydrogen peroxide (Figure 4). Hydrogen peroxide, an inherent product of the pneumococci's anaerobism is linked to inflammation, and an increase in PARr on the hosts epithelial cells, a receptor necessary for invasion (Bogaert et al., 2004).The precise role of the major autolysin LytA remains unclear, however its ability to cleave lactyl-amide bonds between peptide and glycan strands in the cell wall means it likely has important consequences on cell morphology, and growth (Tomasz et al., 1988). LytA can also cause

lysis of pneumococci, leading to the release of additional virulence factors such as pneumolysin (Mellroth et al., 2012). Pneumolysin (Ply) is a cholesterol dependent cytolysin. It binds membrane bound cholesterol on the surface of eukaryotic cells, forming 400-Å pores resulting in cell lysis (Alouf, 2000, Shak et al., 2013). Pneumolysin consequently activates complement during its attack on host cells (Kadioglu et al., 2008).

Invasion is preceded by the local generation of IL-1, tumour necrosis factor (TNF), and other cytokines, which results in an inflammatory cascade, causing changes in the expression of immune receptors present on the epithelial cells. PAFr is up-regulated during this process (Tuomanen, 1997, Bogaert et al., 2004).

Following up-regulation of PAFr, phosphocholine, present in the pneumococcal cell wall, shows an increased affinity for this epithelial-cell receptor (Bogaert et al., 2004). Choline binding protein A (CbpA), alternatively named PspC or SpsA, also shows an increased affinity towards immobilised sialic acid and lacto-N-neotreatose residues on the epithelial cells surface. In addition CbpA binds to the human secretory component part of polymeric immunoglobulin receptors (PIgR) expressed on the surface of epithelial cells (Kadioglu et al., 2008). This greater adherence to the epithelial cells results in an increase in migration of pneumococcal cells across the mucosal barrier- transcytosis (Bogaert et al., 2004).

A CbpA variant, H-binding Inhibitor of Complement (Hic), covalently bound to the cells surface by a LPXTG motif, further acts to subjugate complement-mediated clearance (Janulczyk et al., 2000). This is achieved through its ability to bind host factor H component, an alternative complement pathway inhibitor (Jarva et al., 2002, Neeleman et al., 1999).

Other related proteins such as PspA are likewise found to inhibit complement activation and opsonisation by C3 (Abeyta, 1999, Tu et al., 1999, Ren et al., 2003, Ren et al., 2004).

Immunoglobulin defences are subjugated during this processs through the expression of a zinc dependent proteinase (IgA1prot), which cleaves immunoglobulin A1 at the hinge region (Kilian et al., 1988, Wani et al., 1996)- IgA being the most abundant immunoglobulin present in mucosal secretions (Fagarasan and Honjo 2003).

Internalisation corresponds to the binding of pneumococcal ChoP to PAFr (Cundell et al., 1995, Bogaert et al., 2004). Transcellular migration of living bacteria through the respiratory epithelium and vascular endothelium ultimately leads to invasive disease (Bogaert et al., 2004).

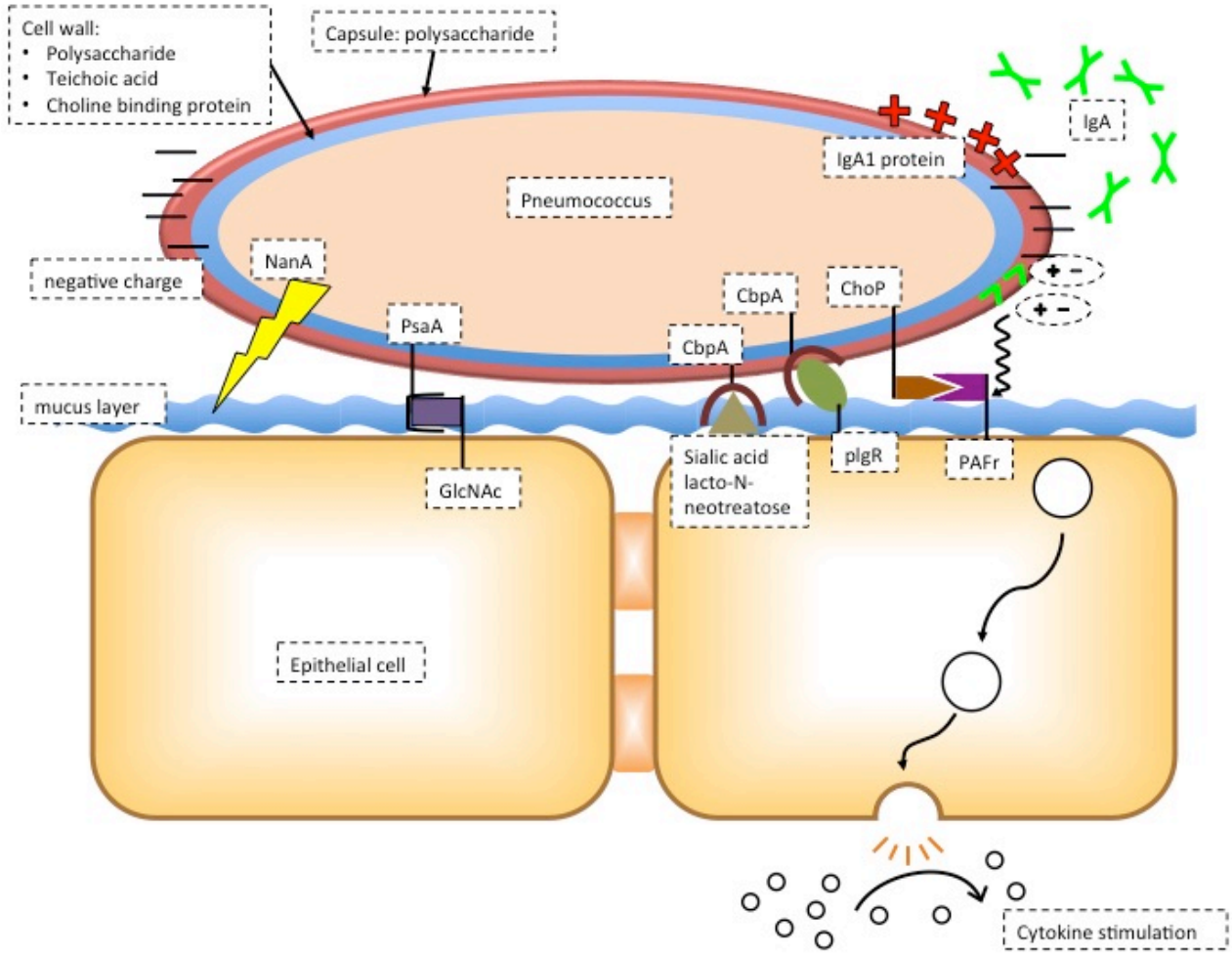


Figure 4: Schematic showing the mode by which pneumococci are thought to attach and invade the host's epithelial cells. GlcNAc receptors on the epithelial surface are exposed as NanA reduces the viscosity of overlying mucus. GlcNAc can then interact with pneumococcal surface proteins such as PsaA. PAFr becomes upregulated on the epithelial cells surface in response to cytokine stimulation. Increased affinities of ChoP for PAFr and CbpA for immobilised sialic acid and lacto-N-neotreatose, in addition to the binding of CbpA to PlgR aids transcytosis. Binding IgA is subjugated by the cleavage action of pneumococcal IgA1 protease (from Bogaert et al., 2004).

1.3.6 Other virulence components

The pneumococcus has also been found to possess virulence components that enhance invasion, but are not ubiquitous to all isolates. For example, invasive disease is more often associated with the presence of a pneumococcal serine-rich repeat protein (psrP) pathogenicity island (Obert et al., 2006). PsrP is found to bind keratin 10 (K10), which is present on the surface of lung cells, but not on nasopharyngeal cells (Shivshankar et al., 2009). The gene, pneumococcal collagen like protein A (PclA), was similarly found to be present in ~40% of clinical isolates in one study, and also acts to enhance adherence and invasion of host cells (Paterson et al., 2008).

In addition to virulence components, viral co-infection can also facilitate invasive pneumococcal disease. Influenza virus co-infection is closely linked to increased pneumococcal mortality (Farr, 1885, Simonsen et al., 2000, Morens et al., 2008). Co-infectione is once again linked to upregulation in PAF receptor on the host's epithelial cells facilitating pneumococcal attachment and transcytosis (McCullers and Rehg, 2002). More recent studies have linked stimulation of type 1 interferons (IFN), which reduces CCL2 chemokine production, to reduced clearance. This results from impaired macrophage recruitment to the infection site by CCL2 (Nakamura et al., 2011).

1.3.7 Serotype and Disease

The capsule region is highly variable, with over 90 pneumococcal serotypes currently recognised. Such variability is thought to have emerged as a result of heavy immune selection favouring diversification (Lipsitch and O'Hagan, 2007).

With the exception of serotype 3 and 37, the capsule genes are encoded at a single locus, between glucan 1,6-α-glucosidase gene, *dexB,* and oligopeptide ABC transporter, *aliA* (although neither *dexB* or *aliA* are involved in capsule biosynthesis)(Figure 5)(Kolkman et al., 1998, Garcia et al., 2000). The locus consists of four constant genes (*capABCD*) at the 5' end, and a serotype specific region, which is variable in gene number depending on capsule complexity (Bentley et al., 2006). An additional set of 3' flanking genes are present in cases where the capsule contains glucoronic acid (e.g. 2, and 8) or rhamnose (1, 2, 6B, 18C, 19F, and 19A) (Waite et al., 2001). Capsule synthesis, excluding serotypes 3 and 37, occurs via the Wzx/Wzy-dependent pathway, through a repetitive process of polymerisation and translocation of polysaccharide subunits across the cell membrane via the Wzd/Wze complex (James and Yother, 2012). Transfer of this repeat unit by the Wzx flippase to the outer surface of the membrane is followed by its subsequent polymerisation through Wzy. The resulting polysaccharide will either remain associated with the membrane, be released, or be attached to the peptidoglycan cell wall (James and Yother, 2012).

The capsules of serotype 3 and 37 are comparatively simple in structure. The serotype 3 capsule is composed of repeating cellobiuronic acid units (D-glucuronic acid β1,4 linked to glucose)(Hotchkiss and Goebel, 1937, Heidelberger and Hobby, 1942), whereas repeating sophorosyl units form the serotype 37 capsule (Llull et al., 1999). In both serotypes (3 and 37), capsule biosynthesis proceeds through processive transferase activity, compared to the formation of lipid-linked intermediates in other pneumococcal types (Arrecubieta et al., 1996, Llull et al., 1999). The serotype 3 capsule is encoded by *cap3ABCD* and *galU* and *pgm* genes. *Cap3ABCD* occur in a single cluster in the normal capsule locus, although only mutations in *cap3A* and *cap3B* result in loss of capsule, suggesting gene homologues exist elsewhere in the genome for *cap3C* and *cap3D* (Dillard and Yother, 1994, Arrecubieta et al., 1996). The existence of homologues therefore makes this serotype more resilient to capsule loss by mutation or gene disruption than other pneumococcal serotypes.

UDP-glucose precursors are synthesised by Cap3CD, GalU and Pgm enzymes, and may be processed further by Cap3A (UDP-Glucose dehydrogenase) to form UDP-glucoronic acid. Cap3B (Polysaccharide synthase) then uncouples gluconic acid or glucose from the UDP sugar group, and forms the cellobiuronic acid units of the polysaccharide chain and connects these together by a β1,3 linkage (Arrecubieta et al., 1996). Capsule synthesis in the 37 serotype is driven by the β-glycosyltransferase Tts enzyme, which catalyses the glucosidic linkages (1,2 and 1,3) necessary for capsule synthesis in this serotype (Llull et al., 1999)

The capsule type (serotype) carried is often associated with differing durations of nasopharyngeal carriage, potentials to cause disease, virulence, and, often due to a shared ancestry, antimicrobial susceptibility (Fedson et al., 2011, Simell et al., 2012).

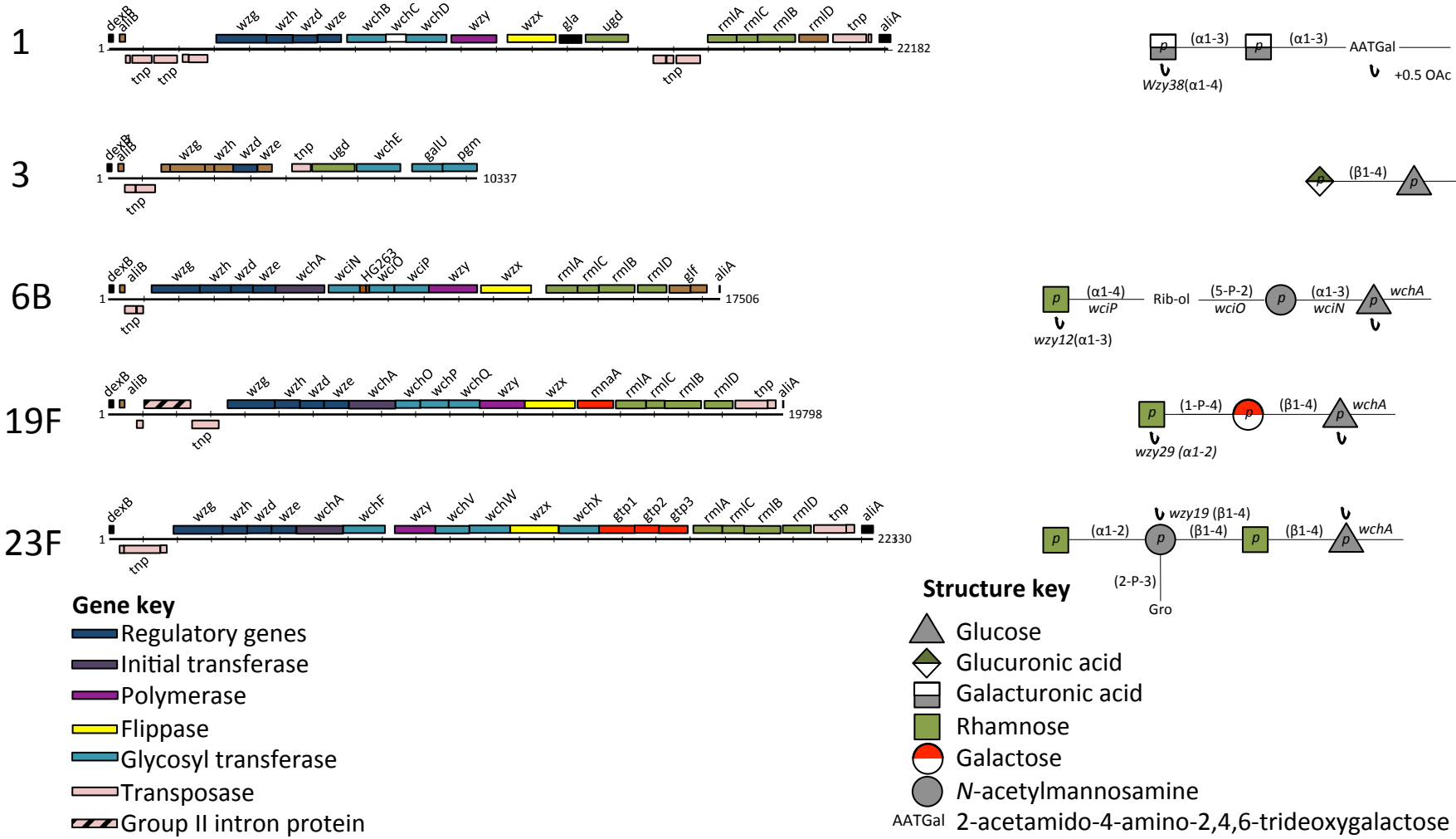


Figure 5: Variation in the capsule structure of three important pneumococcal serotypes. The capsule locus is usually flanked by the genes *dexB* and *aliA*. There follow a set of regulatory genes, before the more variable polymerase, flippase and transeferase units. The locus is then commonly concluded with a set of relatively conserved flanking genes (from Bentley et al., 2006).

Invasive disease is frequently associated with serotypes 1 and 7F (Fedson et al., 2011), and is rarely associated with long-term carriage (Gray et al., 1980). In contrast, serotypes 3, 6A, 6B, 8, 19F and 23F are identified as having low invasive potentials- predominantly causing disease among high risk rather than otherwise healthy individuals (Fedson et al., 2011). The most virulent disease is similarly associated with serotype 3, although this serotype usually exhibits a low invasiveness potential (Scott et al., 1996, Nielsen and Henrichsen, 1992, Ahl et al., 2013).

The association of disease with particular serotypes has lead to the estimation that 70% of IPD worldwide is associated with only 20 of the ~90 circulating serotypes (Kyaw et al., 2006). Whilst such figures indicate that vaccines that target such serotypes should help to largely combat disease, such prevalences are subject to seasonal, and spatial variations, as well as variation between age groups (Bogaert et al., 2011).

In Malawi, serotype 1 accounts for 20.5% of IPD, although being rarely found in carriage. In contrast serotypes 2 and 14 are more frequent causes of IPD in the UK (Brueggemann et al., 2003, Everett et al., 2012), and serotypes 19F and 23F in South-East Asia (Jauneikaite et al., 2012). Vaccination has had a profound effect on serotype distribution in those countries where it has been introduced, with substantial reductions in vaccine type serotypes in the years following vaccination (Gladstone et al., 2015), which are often replaced by other circulating serotypes (Weinberger et al., 2011).

Historically, the greater association of multidrug resistance (typically strains resistant to three or more antibiotics) with particular serotypes (Hackel et al., 2013) appears to have resulted from the clonal expansion of a limited number of pneumococcal lineages (Jabes et al., 1989, Munoz et al., 1991).

1.3.8 PMEN lineages

Since the late 1980s the importance of clonal decent in pneumococcal drug resistance has become increasingly recognised (Markiewicz and Tomasz, 1989, Jabes et al., 1989). Pulsed field gel electrophoresis (PFGE)(Schwartz and Cantor, 1984), and genetic sequencing techniques allowed for more thorough investigations to be carried out into the penicillin-binding-proteins, the primary targets of beta-lactam antibiotics (Hakenbeck et al., 1980, Zighelboim and Tomasz, 1980). Such studies suggested that multidrug resistance had been acquired multiple times, by pneumococci belonging to different clonal lineages (Jabes et al., 1989). It furthermore appeared that resistance-encoding penicillin-binding proteins (*pbp*) were initially acquired by genetic transfer of *pbp* alleles from other co-colonising species present in the nasopharyngeal niche (Dowson et al., 1990, Laible et al., 1991). Pneumococcal clones, that had acquired resistance, were then thought to have spread globally, carrying multi-drug resistance (MDR: resistance to beta-lactams and at least two other antibiotic classes) with them (Jabes et al., 1989).

Clinical isolates carrying a serotype 23F capsule and carrying penicillin, tetracycline and chloramphenicol resistance were frequently isolated in Spain during the early 1980s (Latorre et al., 1985). Serotype 23F isolates were similarly more frequently associated with MDR than any other serotype in the US (George et al., 1987). Restriction mapping of fragments of *pbp* genes, in conjunction with PFGE analysis of PBP proteins, confirmed the common ancestry of the isolates from these two countries (Munoz et al., 1991). This lineage was not only found to have carried MDR between these two countries, but was also identified as having acted as a source of resistance alleles for other pneumococcal lineages (Coffey et al., 1991).

Further investigations revealed a number of clonal global pneumococcal groups existed (Munoz et al., 1992, McDougal et al., 1992, Coffey et al., 1996, Smith and Klugman, 1997, Shi et al., 1998). Concerns about

treatment breakdown as a result of the global spread of MDR clones lead to the establishment of the Pneumococcal Molecular Epidemiology Network (PMEN) in 1997. The aim of PMEN was to aid global surveillance of clinically important *S. pneumoniae* and develop a standardised nomenclature for the classification of these clones. Such clones must have a wide geographic distribution (isolated from a minimum of two continents), and be resistant to one or more antibiotic in wide clinical usage, or be globally prevalent in disease. Enright and Spratt's Multi Locus Sequence Typing (MLST) scheme became an important method for classifying clones at this time. The method is based on identifying ~7 highly conserved, "house-keeping", genes within the species. Sequence analysis of ~450bp fragments of these genes can then be used to identify ancestry between strains, which are assigned to a sequence type (Enright and Spratt, 1998).

The serotype 23F MDR clone first described in Spain and the US (Munoz et al., 1991) was the first to be included in the PMEN surveillance programme. Assigned the name PMEN1, it is generally characterised as being of sequence type 81 and serotype 23F (McGee et al., 2001). This lineage is thought to have originated in Barcelona, Spain, in the late 1970s, before rapidly spreading to Africa, Asia and America (Munoz et al., 1991, Klugman et al., 1994, Parry et al., 2002). Other PMEN clones were identified, PMEN2 (6B-2), and PMEN3 (9V-3) both originating in Spain being later added to the list, and the number of clones under surveillance currently totals 43 (McGee et al., 2001). The spread of MDR is not only a problem for developed countries, but for resource poor countries, such as Malawi. Here treatment options are limited, and although antibiotic resistance is currently low, recent sequencing studies have revealed the emergence of PMEN clones within the local population (Everett et al., 2012).

1.3.9 The success of MDR clones

The prevalence of MDR clones globally represents a paradigm. Antibiotic usage provides a selective pressure, favouring the acquisition of resistant mutations, and favouring the replacement of antibiotic susceptible clones by drug resistant lineages (Cohen et al., 1997). However, in the absence of antibiotic pressure empirical evidence suggests that such mutations should be deleterious (Rudolf et al., 2011). A fitness cost in drug-resistant isolates would be expected to translate into a reduction in prevalence and extinction from the gene pool in the absence of antibiotic therapy (Trzcinski et al., 2006).

In reality, resistant pneumococcal strains were frequently isolated prior to PCV7 (Appelbaum, 1987). Resistance is similarly largely conserved globally throughout clonal lineages, despite different, country specific, antibiotic regimes (Croucher et al., 2011, Orio et al., 2011).

Biological fitness is measured in terms of the ability to transmit and colonise hosts (Trzcinski et al., 2006). This causes difficulties when trying to assess the fitness costs of mutations empirically, so that the costs associated with putative resistance SNPs must be assessed via indirect means, such as the colonisation rates of these bacteria (Trzcinski et al., 2006). Such studies have suggested that resistance appears to show stability even in the absence of antibiotic, with normal rates of growth and physiology under *in vitro* cultivation (Tomasz, 1997).

A recent assessment of this paradigm was carried out by Orio and colleagues (2011). This study used strains belonging from the international Spain^{9V}-3 clone, isolated from invasive infections of paediatric patients, which were assessed using the beta-lactams, piperacillin, cefotaxime, and penicillin (Orio et al., 2011). During competition experiments with wild-type strains, only those mutations associated with *pbp2b* were found to confer significant fitness costs to resistant isolates (Orio et al., 2011). However, it appeared that such

deleterious mutations were entirely compensated for by mutations acquired by *pbp2x* and *pbp1a* (Orio et al., 2011). Although compensation was not demonstrated for all strains, this finding suggests that such mechanisms are likely to have been important for the spread, long-term maintenance and success of MDR pneumococcal lineages (Orio et al., 2011).

1.4 Antibiotics Overview

A variety of antibiotics are currently available for treating bacterial infections (Figure 6). These vary in terms of their target sites, bacterial specificities, modes of action, and methods of administration. Although beta-lactam resistance is now recognised globally, these remain the drugs of choice for treating pneumococcal diseases, being widely used in both developing and developed countries. They exert a powerful bactericidal activity by targeting cell wall synthesis enzymes. In addition, their widespread availability and relative low cost has allowed them to be used globally. In Malawi, the beta-lactams, penicillin and ceftriaxone remain first-line pneumococcal treatments. For this reason, the study presented here focussed on the beta-lactam class of antibiotics.

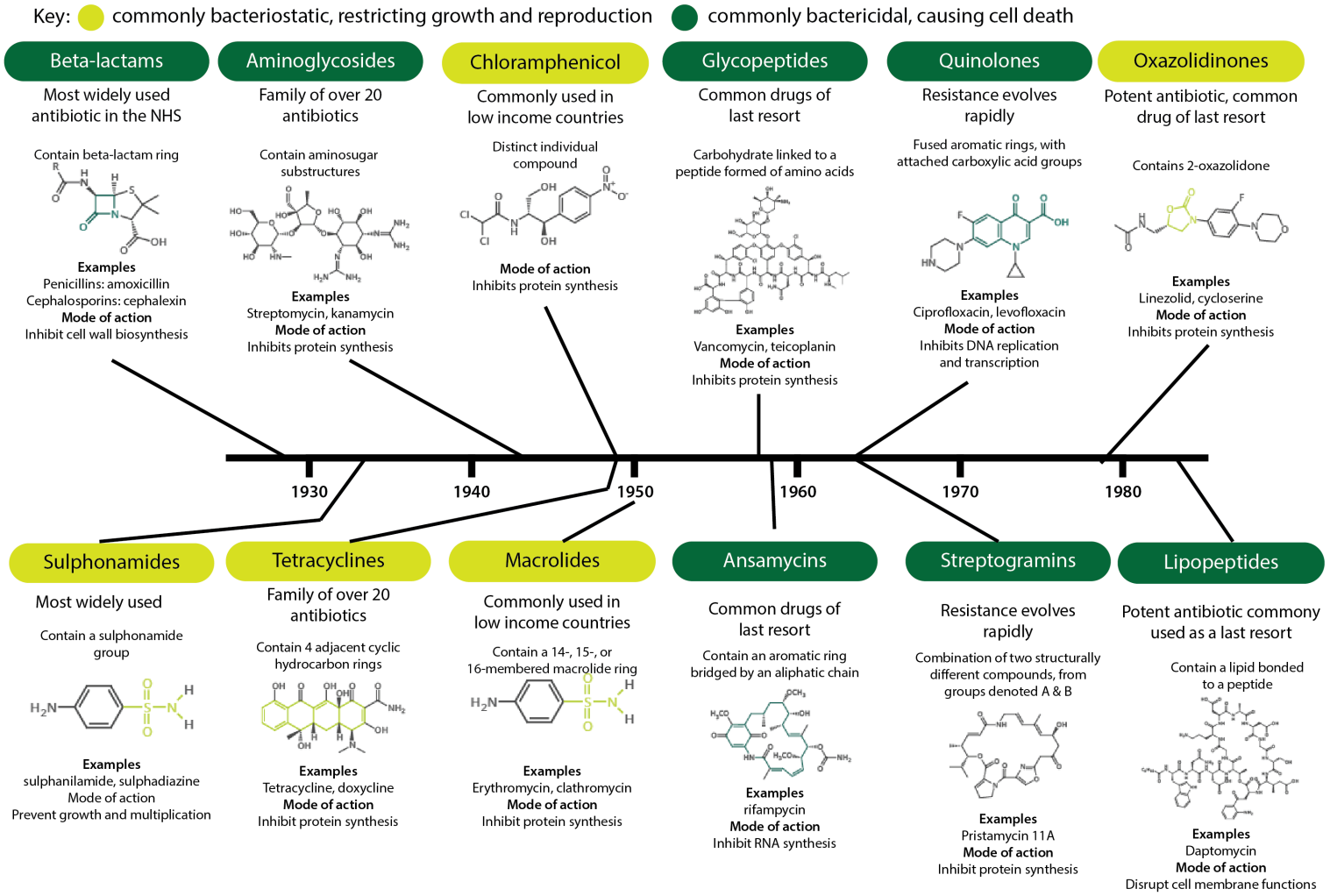


Figure 6: Diagram showing the dates of discovery of different antibiotic classes and their general structures (Compound Interest, 2014).

1.5 The Beta-lactams

1.5.1 Cell wall synthesis

The cell wall is composed of teichoic acids and peptidoglycan (murein) chains, composed themselves of numerous glycan chains cross-linked by peptide bonds (Ruiz, 2008). The wall acts as an anchor point for a number of cell wall associated surface proteins (Bogaert et al., 2004), regulates molecular uptake and secretion, maintains tensile strength and influences bacterial form (Ruiz, 2008, de Kruijff et al., 2008). Beta-lactams target the bacterial cell wall synthesis pathway, resulting in wall rupture, and cell death.

Wall precursor synthesis is often divided into two stages: the first results in the synthesis of the N-acetylmuramylpentapeptide, (MurNAc pentapeptide) which occurs in the cytoplasm, whilst the second step linking the MurNAc pentapeptide to a lipid carrier, occurs within the membrane prior to integration of these monomeric building blocks into the developing cell wall (de Kruijff et al., 2008).

The first two reactions of precursor biosynthesis are catalysed by the enzymes MurA and MurB respectively and result in the conversion of the N-acetylglucosamine (GlcNAc) into N-acetylmuramic acid (MurNAc) for further processing (Macheboeuf et al., 2006). There follows the stepwise addition of peptide groups to the exposed D-lactyl residue of the MurNAc molecule, by enzymes MurC to MurF (Figure 7)(El Zoeiby et al., 2003).

In the second stage, the enzyme MraY catalyses the attachment to the membrane bound undecaprenyl phosphate group (bactoprenol), resulting in the formation of membrane bound lipid I (Lecercle et al., 2010). A GlcNAc is then attached to the MurNAc head group of lipid I by the enzyme MurG resulting in the formation of lipid II (Figure 7)(Macheboeuf et al., 2006).

In gram-positive bacteria such as *S. pneumoniae* lipid II may undergo further processing. Short dipeptides such as serine-alanine, or additional di-alanine groups may be attached to the L-lysyl group in the third position in the MurNAc pentapeptide chain (Lloyd et al., 2008). This process requires two enzymes, MurM which uses either Ala-tRNA^{Ala} to append Ala, or Ser-tRNA^{Ser} or preferentially the mischarged Ser-tRNA^{Ala} to append Ser to this stem pentapeptide, followed by the enzyme MurN which then attaches Ala to the previous amino acid, again via Ala-tRNA^{Ala} group completing the septapeptide (Lloyd et al., 2008).

Once secondary processing is complete, a flippase enzyme within the membrane transports the lipid II molecule from the cytoplasm onto the periplasmic side of the cells membrane (Ruiz, 2008). Glycosyltransferase enzymes then catalyse the joining together of consecutive lipid II monomers forming glycan chains, and resulting in the release of the bactoprenol carrier molecule which is recycled back into the membrane (Ruiz, 2008, Breukink and de Kruijff, 2006). A series of transpeptidation and carboxypeptidation reactions are then necessary to join together the newly synthesised glycan chains thus yielding peptidoglycan (Ruiz, 2008).

The formation of cross-links between glycan chains involves the pentapeptide stem groups (Ruiz, 2008). This final step involves a donor and acceptor strand which interact so that the terminal D-alanyl-D-alanine group of the acceptor strand is cleaved, allowing a cross-linking peptide bond to be formed between the third position L-Lysine group of the acceptor and the fourth position D-alanine of the donor stem pentapeptide chains (Macheboeuf et al., 2006)(fig. 1).

It is this final set of reactions that are catalysed by the penicillin binding proteins, which all share a common DD-peptidase activity, (DD-transpeptidase, DD-carboxypeptidase or DD-endopeptidase) (Macheboeuf et al., 2006).

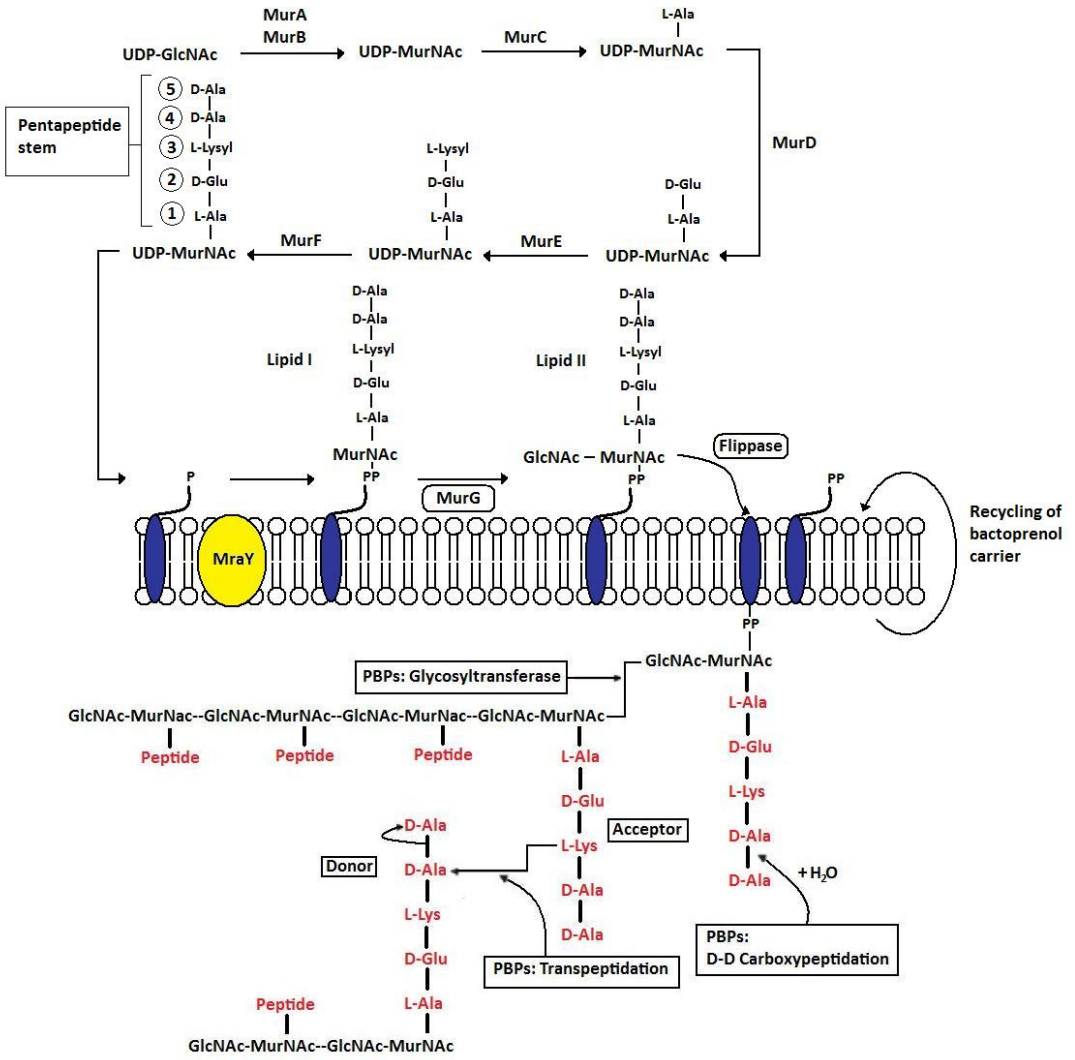


Figure 7: Sequence of events leading to the synthesis of peptidoglycan. Uridine diphosphate acts as an energy source for the initial reactions leading to the formation of MurNAc and the pentapeptide stem. The bactoprenol carrier (purple) and MraY (yellow) catalyse the synthesis of lipid I. The MurG enzyme then attaches GlcNAc thus forming lipid II. This molecule is then "flipped" to the external side of the membrane by an unknown flippase enzyme where synthesis is completed by membrane bound penicillin-binding-proteins (adapted from Macheboeuf et al., 2006).

1.5.2 Penicillin Binding Proteins

PBPs are divided into three classes according to the catalytic domains that they possess (Figure 8)(Sauvage et al., 2008). The high molecular weight class A PBPs catalyse both glycosyltransferase and transpeptidase reactions and hence are bifunctional. Class B PBPs catalyse transpeptidase reactions alone and as such are monofunctional. The low molecular weight, D,D-carboxypeptidase PBPs are also monofunctional, capable of carboxypeptidation reactions alone. All three classes of PBP share an affinity towards the D-alanyl-D-alanine group, the substrate of

the transpeptidase step (Macheboeuf et al., 2006). High affinity β-lactams compete with this substrate for the active site of these PBPs, preventing crosslinking between glycan chains and leading to the weakening of the bacterial cell wall (Sauvage et al., 2008).

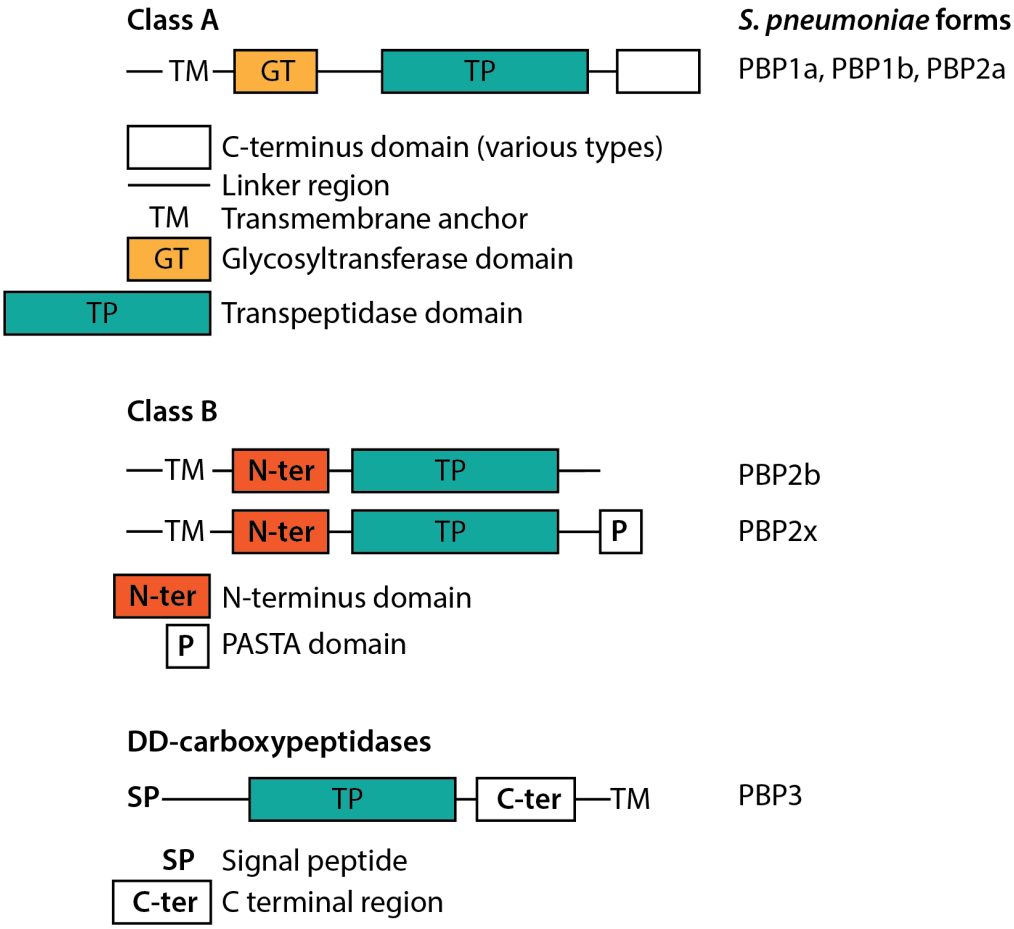


Figure 8: The composition of penicillin binding protein domains. Additional domains to the glycosyltransferase and transpeptidase domains are shown. PBP2x possesses a PASTA (PBP and Serine/Threonine kinase Associated domain) domain, capable of binding the beta-lactam stem, suggesting it plays a role in sensing D-alanyl D-alanine (Yeats et al., 2002). N-terminus and C-terminus domains play a role in anchoring of the PBP to the cell membrane (modified from Macheboeuf et al., 2006).

The carboxypeptidation and transpeptidation reactions catalysed by PBPs occur via a three-step pathway. The first step yields the rapid and reversible formation of a non-covalent Henri-Michaelis complex through the action of enzyme and the stem pentapeptides of the donor strand (Lim and Strynadka, 2002). This is followed by attack of the active serine residue, located in the PBP active site, on the substrates carbonyl carbon atom located at the C-terminal of the D-alanyl-D-alanine peptide bond

(Figure 9)(Sauvage et al., 2008). This leads to the formation of the acyl-enzyme intermediate, and also leads to the concomitant cleavage of the C-terminal D-alanine group, a process termed acylation (Macheboeuf et al., 2006).

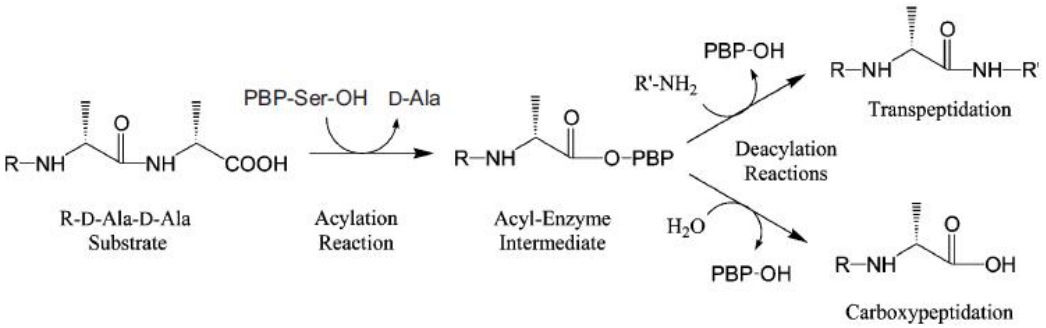


Figure 9: Steps two and three in the synthesis of peptidoglycan, demonstrating serine attack (acylation), and the two possible outcomes of this reaction, bond formation and transpeptidation or hydrolysis and carboxypeptidation (Nicola et al., 2005).

Acylation involves the removal of a proton from the active serine (S) by the deprotonated lysine residue (K) located in the highly conserved SxxK (S: serine, k: lysine) motif (Figure 10) (Bobba and Gutheil, 2011). This proton is then donated to the amine group of the cleaved D-alanine group via the serine molecule of a second SxN (N: asparagine) motif, forming a PBP-substrate compound (Sauvage et al., 2008). The process is completed (deacylation) with either the hydrolysis and release of a shortened peptide (carboxypeptidation), or the formation of a cross-link between the donor strand and a second peptidoglycan pentapeptide stem group of an acceptor molecule, a process termed transpeptidation (Nicola et al., 2005). Deacylation requires the removal of a proton from an acceptor (an amino-group from an adjacent pentapeptide chain in transpeptidation, and a water molecule in carboxypeptidation) (Sauvage et al., 2008). This allows the "activated" acceptor molecule to attack the ester bond between the acyl-enzyme complex, at the carbonyl carbon atom, releasing the substrate from the PBP active site (Sauvage et al., 2008). The PBP active site may therefore be viewed as a "double lysine-

serine system", responsible for both acylation and deacylation (Sauvage et al., 2008).

The PBP binding domain exists at the interface between these two subdomains: one, an alpa-helical domain, and the other a five-stranded β-sheet, overlaid by three α-helices (Figure 10). Here, a degree of flexibility between these subdomains is thought to influence the binding capabilities of some PBPs (Lim and Strynadka, 2002). Three amino acid motifs are highly conserved (Denapaite et al., 2007). Located at the start of the α2 helix is the active serine group, which is followed by a lysine group, together forming the previously mentioned SxxK motif. A SxN motif is formed through a loop formed between α-helices four and five, whilst a third KT(S)G (T: Threonine) motif results from the interaction between four residues on β3-sheet. An additional conserved glycine residue is also located towards the rear of the active site (Figure 10).

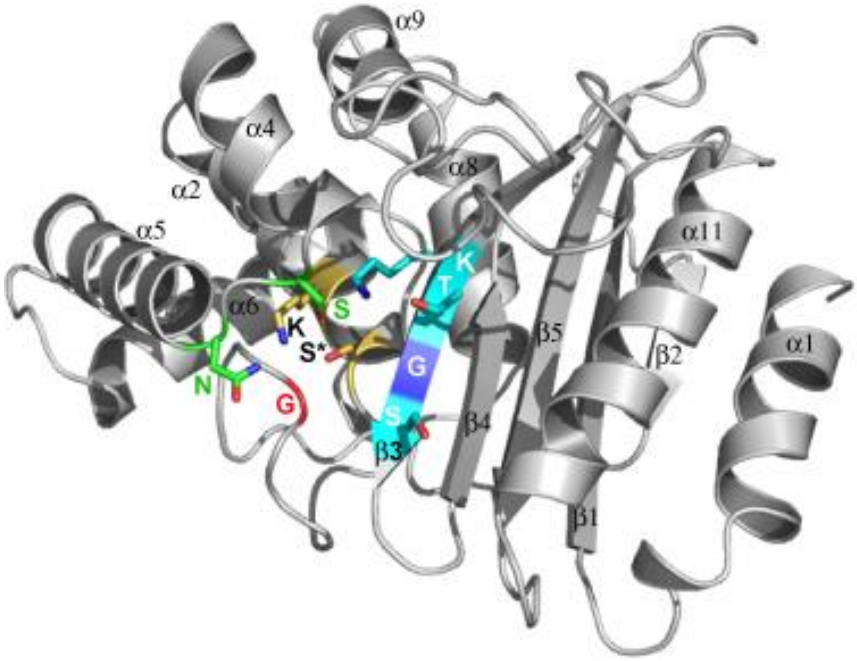


Figure 10: Tertiary structure of the penicillin-binding-domain. Alpha helices 1 to 9 are labelled, in addition to beta-sheets 1 to 5. The SxxK motif is highlighted in blue and cyan, the SxN motif is shown in green, whilst the third KT(S)G motif is shown in yellow. The conserved glycine residue of this motif, which is located at the rear of the active site, is labelled in red. Additional colours: blue represents exposed amide groups, whilst red illustrates exposed hydroxyl groups (from Sauvage et al., 2008).

To date, knowledge regarding the interaction between substrate and active site is largely based on studies of low molecular mass (LMM) PBPs from different bacterial species, interacting with substrate mimics, such as D-α-amino-pimelyl-ε-D-alanine (Figure 11) a molecule that mimics the C-terminal end of Glc-NAc-MurNAc stem pentapeptide (Denapaite et al., 2007, Sauvage et al., 2007).

Current models suggest that the penultimate D-alanine group of the substrate is tightly held within the active site following covalently binding to the enzymes active serine (Figure 11)(Sauvage et al., 2007). Further "wedging" of the amide group between the asparagine molecule, carried by the second SxN motif, and the backbone of the β3-strand that lines the active site helps to manipulate the substrates stem group within the active site. The carbonyl group of the penultimate D-alanine group is also found to fit into the oxyanion hole, its methyl group being located within a hydrophobic pocket created through the conserved glycine residue located at the back of the active site (Sauvage et al., 2008).

Two hydroxyl groups found on the third KT(S)G motif are orientated towards the carboxyl group of the leaving D-alanine peptide. Thus it appears that the asparagine of the second motif in combination with these two hydroxyl groups carried by the third motif are highly important to the correct positioning of the substrate within the active site. The glycine carried by the third motif appears to prevent steric hindrance as larger molecules would prevent substrate associating fully with the active site. A glycine molecule at the rear of the active site also appears to play a role in increasing binding specificity of the PBPs towards the penultimate D-alanine group, as larger molecules could lead to steric hindrance of the active site (Sauvage et al., 2008).

Studies using x-ray analysis have demonstrated that β-lactam antibiotics covalently bind to the active site serine forming a relatively stable intermediate within the PBP transpeptidase domain, blocking the active

site from further catalysis. The antibiotic is thought to fit this site through the insertion of an amide group side-chain between the asparagine of the second KT(S)G motif and the backbone of the β-3 strand. This in addition to hydrogen bonds formed between antibiotic and enzyme are thought to ensure a tight fit between active site and substrate (Sauvage et al., 2008).

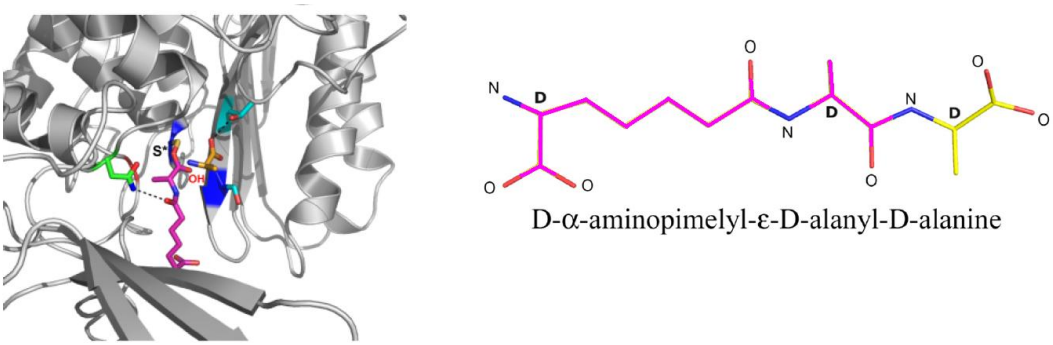


Figure 11: PBP binding domain (*Bacillus subtilis* PBP4a), showing the interaction with the substrate mimic, D-α-aminopimelyl-ε-D-alanyl-D-alanine. The substrate is covalently bound to the active serine residue, and the cleaved terminal D-alanine group (yellow) is held via hydrogen bonds (dashed lines in 3D figure) to residues located on β-3. The red "OH" indicates the location of the oxyanion hole, formed by the two backbone amine groups (dark blue). Note, pneumococci possess a lysine residue rather than diaminopimelic acid (DAP) at the position 3 acceptor (from Sauvage et al., 2008).

1.5.3 PBPs in pneumococci

The HMM Class A PBPs are essential for growth in many bacteria. Of the six PBPs identified in *S. pneumoniae* (Maurer et al., 2008, Reinert, 2009), three of these belong to Class A; PBP1a, PBP2a, PBP1b (Hoskins et al., 1999). Gene inactivation studies indicate that either PBP1a or PBP2a must be present for growth *in vitro,* whereas individually the other PBPs are dispensable (Hoskins et al., 1999). Class A PBPs are composed of a cytoplasmic N-terminal domain coupled to a membranous region, together forming a transmembrane anchor (Figure 8) (Di Guilmi et al., 1998). There follows a glycosyltransferase domain, joined to a further transpeptidase domain and the molecule is completed with a Ser- and Asn-rich C-terminal extension (Di Guilmi et al., 1998). Variants of this general structure are further described by Sauvage and colleagues (2008). *S. pneumoniae* possesses two Class B PBPs; PBP2x, and PBP2b. The functions of these two PBPs differs slightly, with PBP2x being involved in cell division, whilst PBP2b plays a role in cell septation. The Class B PBPs are comprised of an N-terminal domain joined to a C-

terminal transpeptidase domain (Figure 8). One difference between class II and class I proteins is that the residue that follows the KT(S)G motif of the transpeptidase domain is typically an alanine in Class B PBPs, being a threonine or serine in class A PBPs (Sauvage et al., 2008). A limited number of penicillin resistant PBPs, including PBP2x, possess an additional C-terminal domain resulting from one or two repeating units, identified as Penicillin-binding protein And Serine/Threonine kinase Associated domains (PASTA). The PASTA domain consists of three β-strands and one α-helix and is thought to play a role in the binding of peptidoglycan to such molecules (Figure 8) (Yeats et al., 2002)

The third low molecular mass D,D-carboxypeptidase, of which *S. pneumoniae* possesses one, PBP3 (*pbp3*), although not essential *in vitro,* is thought to be important for organizing growth and division processes (Morlot et al., 2005, Sauvage et al., 2008). This protein is composed of a short signal peptide attached to a transpeptidase region terminating in a transmembrane domain (Figure 8) (Macheboeuf et al., 2006).

All six PBPs have been implicated in β-lactam resistance to varying degrees (Krauss and Hakenbeck, 1997, Hakenbeck et al., 1998). However high-level β-lactam resistance is most often associated with a number of complementary mutations occurring among the *pbp*1a, *pbp*2b and *pbp*2x genes. Furthermore, pbp genes frequently contain sections of sequence that have non-pneumococcal origins, termed mosaicism (Dowson et al., 1993, Coffey et al., 1995, Orio et al., 2011).

1.5.3 Beta-lactam resistance

The current model for penicillin resistance suggests that mutations first arise within the *pbp2x* gene. Modification of *pbp2b* follows, resulting in further decreases in susceptibility, before final modifications in the *pbp1a* gene results in a between 10 and 20 fold increase in resistance (Barcus et al., 1995, Smith and Klugman, 1998, Pernot et al., 2004). Such studies reveal that mutations are required to occur in a set order, those

mutations occurring prior to their position within this pathway being essentially cryptic, with no phenotypic effects evident.

Beta-lactams exhibit varying affinities towards different PBP target sites, for example piperacillin binds all PBPs with high affinity, whereas cefotaxime does not appear to bind PBP2b (Grebe and Hakenbeck, 1996). Similarly, *pbp2x* is the first gene to be modified in the development of cefotaxime resistance, whereas *pbp2b* is the first gene to be modified in the development of piperacillin resistance (Grebe and Hakenbeck, 1996). Although different mutations are required within different *pbp* genes for low-affinity variants to develop, for antibiotics such as penicillin, which require a multitude of changes across different *pbp* genes, resistance to other beta-lactams can occur concurrently. Consequently, penicillin resistance is often associated with resistance to other beta-lactams.

A number of studies have attempted to identify the course and stepwise gain of mutations necessary for resistance to develop for specific beta-lactams. However such studies have struggled to provide a clear understanding of this process, complicated by a lack of concurrence between those mutations resulting in a loss of beta-lactam susceptibility *in vitro* with those observed in clinically resistant isolates (Maurer et al., 2008). A major complication is the role that recombination plays in the evolution of this pathogen. With up to 70% of the genome shown to undergo recombination, (Croucher et al., 2011), it is difficult to tease apart those SNPs necessary for resistance to develop, from the high variation present due to recombination. Furthermore, clinical isolates are typically exposed to a diverse selection of antibiotics, and are therefore selected under different conditions than can be achieved in the laboratory (Maurer et al., 2008).

Some widely conserved features, corresponding to a resistant phenotype have however been identified. Coffey and colleagues (1995) studied beta-lactam resistance in PMEN1 strains isolated from the USA. Their results

suggested that high-level resistance to extended spectrum cephalosporin antibiotics resulted from alterations in both PBP1a and PBP2x (Coffey et al., 1995). Sequence alignments between two clinical strains were found to be virtually identical, yet there exhibited a fourfold difference in their resistance profiles with respect to cephalsporins (MICs of 8 and 32 μg cefotaxime/mL respectively). Coffey and colleagues (1995) suggested that this difference was solely attributable to a T to A nucleotide substitution at position 550, occurring in the residue that follows the conserved KTG motif of PBP2x. Whilst this mutation appeared to increase resistance to cephalosporins, it appeared to reduce resistance to penicillin (Coffey et al., 1995). Grebe and Hakenbeck (1996) similarly observed that PBP2x transformants, with a reduced sensitivity to cefotaxime, contained a T to A substitution at position 550. Analysis of other clinical isolates lent further support for the importance of this substitution, where it was found to confer high-level cefotaxime resistance (Maurer et al., 2008).

Some studies have similarly identified non-pbp mechanisms of beta-lactam resistance. Modifications of MurM enzymes associated with the formation of branch chains from the third position lysine residue of the stem pentapeptide have been shown to be essential for the development of high-level penicillin resistance in some settings (Smith and Klugman, 2001, Lloyd et al., 2008). The development of branched peptides is catalysed by the enzymes MurM and MurN, which add short dipeptides, such as serine-alanine or di-alanine to the lysine residue in position three of the pentapeptide (Lloyd et al., 2008). Lloyd and colleagues (2008) identified the presence of an acyl-tRNA molecule (responsible for the esterification of amino acids), which alanylated the peptidoglycan ε-amino group of the stem pentapeptide lysine in preference to its serylation (Lloyd et al., 2008). This led to differing stem peptides between susceptible and resistance isolates indicating that such conformational changes had an important role in the development of reduced beta-lactam sensitivity (Lloyd et al., 2008).

1.6 The acquisition of resistance

1.6.1 Competence

Genetic material acquired by recombination is frequently classified according to whether it is homologous: the donor DNA is highly similar to the hosts genetic material it replaces, or whether it is non-homologous: the donor DNA represents a novel gene, or is highly divergent from the host's DNA it replaces (Vos, 2009). Several routes for the acquisition of this material exist. Transduction describes the uptake of donor DNA instilled by bacteriophage, which subsequently incorporates into the host's genome. Conjugation involves the transfer of DNA by physical contact between donor and recipient cells, commonly driven by genes located on a plasmid. Whereas transformation defines the ability to uptake DNA free in the environment for incorporation into the genome (Vos, 2009). The occurrence of transformation follows the binding of DNA to the cells surface, and is possible with both ssDNA and dsDNA- although the ssDNA has a transforming efficiency of approximately 0.5% compared to that of dsDNA of a similar size (Miao and Guild, 1970).

Griffith and colleagues (1928) demonstrated the existence of a transformable component during their conversion of an avirulent "rough" (unencapsulated) to a virulent "smooth" (encapsulated) pneumococcus, following subcutaneous inoculation of mice with a sub-lethal rough dosage and heat killed smooth form (Griffith, 1928). These transformations of a type 2 (D39/R6) live form by an attenuated dead form of the same, or type 1 or type 3 capsule type, yielded a pneumococcus expressing that respective capsule type (Griffith, 1928). Later Avery and colleagues (1944) demonstrated that DNA was the transforming principle by converting a rough type 2 derivative with material from a type 3 pneumococcus *in vitro,* which was possible on inoculation with cellular filtrate in which proteins, polysaccharides, and lipids had been chemically removed, but was not when DNase had been added.

The ready ability of pneumococci to undergo recombination has lead to it being termed naturally competent. Competence describes the physiological state required to undergo natural transformation, whereas the term natural competence describes the state of being able to take-up naked DNA from the environment and incorporate it into the genome by homologous recombination (Claverys et al., 2006). The ability of pneumococci to readily undergo recombination was not only important for the discovery of this process, but is thought to have been important in the acquisition of resistance genes by this species, by both intra- and inter-species genetic exchange.

Competence in pneumococci was initially hailed as the first example of quorum sensing, as a result of overpopulation within the niche. A subsequent re-definition of this process has however cast doubt on whether the pneumococcus truly undergoes quorum sensing, which requires cell-to-cell signalling, and an assessment of population density, the latter of which pneumococci lack (Claverys et al., 2006). Instead, competence may represent a generalised stress response, which has now been described to occur in several bacterial species. Such stresses may include DNA damaging agents, such as Mitomycin C, and fluoroquinolines, which are found to promote competence in pneumococci (Prudhomme et al., 2006). Similarly, disruption of *purA*, *guaA guaB* (purine biosynthesis genes)(Claverys and Havarstein, 2007), or oxidative stress resulting from antibiotic (Kohanski et al., 2007) or increased mutational burden, are thought to lead to competence induction as a result of detecting genetic damage (Prudhomme et al., 2006, Gagne et al., 2013). Such studies suggest competence is important for the repair of genetic damage, a view contested by Redfield who has proposed that competence could also exist so that genetic material can be used as a source of nutrition, based on experiments conducted using *B. subtilis* and *H. influenzae* (Redfield, 1993a, Redfield, 1993b).

Competence occurs abruptly during the exponential growth of pneumococci, across most cells in the culture simultaneously, being maintained for a period of approximately 15 minutes, before it sharply decays (Claverys et al., 2006). Competence induces the expression of a variety of early, late and, delayed, which are expressed throughout early and late stages of competence, genes (Peterson et al., 2004).

The extracellular presence of Competence Stimulating Peptide (CSP), encoded for by the *comC* gene, appears to activate this process (Pestova et al., 1996). CSP is exported by dedicated ComA, ComB ABC-transporters (ATP-binding cassette) (Hui et al., 1995). ComA functions to cleave-off a double glycine leader peptide from this substrate, before it is exported across the cytoplasmic membrane, through the joint actions of ComA and ComB (Hui et al., 1995, Håvarstein et al., 1997). CSP is detected by the two component regulatory system, ComD/E, which contains a histidine kinase receptor (ComD) and transcription regulation protein (ComE)(Claverys et al., 2006). CSP is thought to bind to the ComD receptor, resulting in its autophosphorylation, passing a phosphoryl group to ComE (transphosphorylation), its cognate response regulator (Pestova et al., 1996). Activated ComE leads to transcription of two identical copies of the *com*X gene; *com*X1, *com*X1 (Peterson et al., 2004). These encode ComX/σ^x, a competence specific, alternative sigma factor, which accumulates to high concentrations in response to CSP (Lee and Morrison, 1999, Luo et al., 2004). ComX, along with RNA polymerase (RNAP) act to express genes necessary for DNA uptake and processing, termed the late com genes (Claverys et al., 2006). ComW is also required for competence, although its function remains to be fully understood (Piotrowski et al., 2009, Tovpeko and Morrison, 2014), but is found to stabilise ComX against proteolysis by ClpE-ClpP protease and stimulates ComX activity (Piotrowski et al., 2009). ComE activity is responsible for a positive feedback loop through the transcriptional activation of *comAB*, which increases pheromone export, and pheromone detection by ComCDE and ComX (Claverys et al., 2006). The *comCDE* operon is strongly upregulated when CSP reaches a critical threshold. ComE and ComX turn off after 5 to 10 minutes; ComX having a short half-life (Luo and Morrison, 2003), whereas ComE appears to be shut off by

an alternative mechanism, perhaps by the accumulation of the unphosphorylated form (Martin et al., 2013). The genes *comAB*, *comCDE*, *comX*, *comW* along with 9 others comprise the "early genes" (Piotrowski et al., 2009). The early genes are preceded by a direct repeat which is bound by ComE (Ween et al., 1999), while the late genes are preceded by a cin-box/combox which is bound by the alternative sigma factor (Claverys et al., 2006).

1.6.2 DNA import

The pneumococcus possesses a system showing genetic organisation analogous to the T2S systems and type IV pilus of gram negative bacteria, and is encoded by seven genes of the *comG* operon: *comGA*, *comGB*, *comGC*, *comGD*, *comGE*, *comGF*, and *comGG* (Muschiol et al., 2015). The putative function of ComGA is a secreted ATPase, which is thought to function in the assembly and retraction of the type IV pilus (Balaban et al., 2014, Laurenceau et al., 2013, Campos et al., 2013, Muschiol et al., 2015). ComGB appears to be located at the base of the pseudopilus, while ComGC appears to be the main constituent of the pseudopilus, although the precise form of the structure is still contested (Balaban et al., 2014, Laurenceau et al., 2013). Debate similarly exists as to whether the pseudopilus is responsible for binding to the incoming DNA and pulling it through the cell wall transferring it to membrane associated ComEA, or whether it simply acts to trap DNA, bringing it close to the cells surface receptors (Mann et al., 2013, Muschiol et al., 2015). The remaining genes, *comGD*, *comGE*, *comGF,* and *comGG* encode a group of minor pilin proteins, which appear necessary for transformation (Balaban et al., 2014), and may be involved in priming the pilus assembly (Cisneros et al., 2012).

Crossing of the cell wall and underlying cytosolic membrane requires a dedicated translocation apparatus, which is specifically assembled for competence. This apparatus comprises ComEA, EndA, ComEC, and ComFA. ComEA is a membrane associated dsDNA receptor (Inamine and

Dubnau, 1995, Provvedi and Dubnau, 1999). ComEC forms a membrane channel for DNA uptake (Draskovic and Dubnau, 2005), whilst EndA acts as a DNA uptake and virulence endonuclease. EndA activity occurs at the midcell where it is recruited, and functions to degrade dsDNA (Bergé et al., 2013).

Based on this system, dsDNA binds to the cells surface, possibly through the pseudopilus, before contacting the membrane associated translocation apparatus. SsDNA is imported in the 3'-5' direction during which breaks are created along its backbone in an apparently random manner (Mejean and Claverys, 1988, Morrison and Guild, 1973, Lacks et al., 1974, Puyet et al., 1990). During this process, degradation occurs in the 5'-3' direction by EndA, breaking down the complementary strand (Mejean and Claverys, 1988, Desai and Morrison, 2007).

Once internalised in the cytoplasm, ssDNA interacts with Single-stranded DNA-binding proteins B and A (SsBA), DNA processing protein A (DprA), Recombinase A (RecA), the methylase, DpnA, and additional soluble proteins in the cytoplasm. Here, DpnA and SsbB function to protect and stabilise the ssDNA (Attaiech et al., 2011, Johnston et al., 2013). DprA then mediates the loading of RecA onto the internalised ssDNA, and helps protect ssDNA from the actions of endo- and exo-nucleases (Mortier-Barriere et al., 2007, Claverys et al., 2009). During this process RecA forms a single helical filament with ATP and the ssDNA. This complex then searches the chromosome for homologous sequence before catalysing the exchange of the complementary strand resulting in a new heteroduplex (Kowalczykowski et al., 1994, Berge et al., 2003, Chen et al., 2008).

Morrison and Guild (Morrison and Guild, 1972) calculated the median length of ssDNA in the cytoplasm to be ~6.6kb, which appears comparable with recent studies (Croucher et al., 2012). Radioactively labelled DNA studies suggest that DNA is incorporated into the genome

within ~15mins of ssDNA appearing in the cytoplasm (Mejean and Claverys, 1984, Berge et al., 2003). The most recent assessments based on whole genome sequencing, predicted recombination events ranged in size from 3 bps to 72,038bp, with a mean occurrence of 6.3kb (Croucher et al., 2011).

Transcriptome studies have suggested that between 105 and 124 CSP-induced genes exist in the pneumococcus (Dagkessamanskaia et al., 2004, Peterson et al., 2004). Guiral and colleagues (2006) found that out of the 91 genes common to both of these studies, 17 early genes, and 60 late genes existed. Only 14 late genes were found important for transformation, their encoding proteins involved in DNA uptake and internal processing (Guiral et al., 2006). Competence has also been found to lead to the expression of genes responsible for allolysis- cell lysis induced by other cells of the same species (Prozorov and Danilenko, 2011). As such some have argued that competence in the pneumococcus be termed an X-state, reflecting the broader phenotype controlled by the genes of the competence signalling pathway (Claverys et al., 2006, Stevens et al., 2011).

CibAB (bacteriocin), murein hydrolases LytA, and CbpD have all been implicated in the involvement of cell lysis and DNA release (Claverys et al., 2007). Other effectors offer protection against such damaging agents, such as CibC protects against bacteriocin. The early gene *comM* similarly prevents LytA and CbpD acting on the host cell (Guiral et al., 2005, Havarstein et al., 2006).

The association between competence and these lytic agents means that cells in the X-state can lyse neighbouring cells that have not entered this state following the external presence of CSP, a process termed fratricide (Claverys et al., 2007).

Two *comC* alleles were identified by Pozzi and colleagues (1996), which encoded CSP variants CSP-1 and CSP-2. Synthetic copies of CSP-1 and CSP-2 were found to induce competence predominantly in strains carrying the complementary *csp* allele (Pozzi et al., 1996). It therefore appears that mixed pherotypes may lead one strain to enter competence over another during mixed colonization. Subsequent fratricidal activity by the competent isolate may offer it a competitive advantage over those cells that have not become competent (Claverys and Havarstein, 2007).

The pneumococcus is known to be highly recombinogenic, which has played an important role in its ability to respond to clinical intervention (Figure 12). A recent study by Hiller and colleagues (2010) identified 23 recombination events during a 7 month period of a strain colonising a single child, which resulted in the replacement of 7.8% of the genome. The per allele recombination to mutation ratio calculated for housekeeping genes of *S. pneumoniae* is estimated at 8.9:1 compared to 4.75:1 in *N. meningitidis* (Feil et al., 2003). Estimates of the relative likelihood that a polymorphism is introduced through recombination rather than point mutation (R/M), although highly variable, also indicate recombination, rather than mutation as being the predominant process- ranging from ~66, based on species wide MLST data, to 7.2 calculated from whole genome sequence data from a single lineage (Feil et al., 2000, Croucher et al., 2011). This contrasts estimates in *S. aureus* where mutation has been estimated to occur 15 times more frequently than recombination (Feil et al., 2003). Whilst the reason for such high recombination rates is currently unknown, mutation appears to play a relatively minor role in the generation of genetic variation within *S. pneumoniae*.

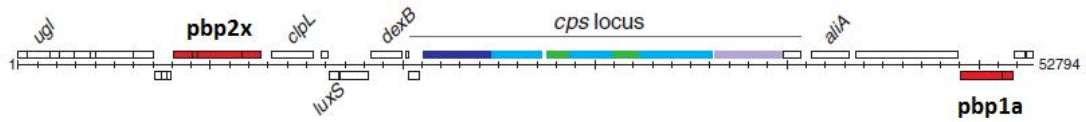


Figure 12: Diagram showing the location of the pbp2x and pbp1a genes (in red) relative to each other and the cps (capsule) locus. Recombination within this region can have particularly important consequences for the development of beta-lactam resistance, and serotype switching (adapted from Croucher et al., 2011).

1.6.3 Recombination efficiency

S. pneumoniae possesses a repair system comprising two Hex proteins, HexA and HexB (Gould et al., 2007). Humbert and colleagues (1995) found that although this mismatch repair system was effective at preventing chromosomal integration with the most and least divergent genetic fragments, it failed to remove intermediately divergent fragments. This appears to result from saturation of the repair system when presented with homeologous (partial divergent) genetic material. Such homeologous material is expected to be representative of genetic material being derived from closely related viridans group Streptococci that share the pneumococci's nasopharyngeal niche (Humbert et al., 1995). For this reason, closely related *S. mitis* and *S. oralis* are thought to have played major roles in the acquisition of drug resistance among pneumococci.

The existence of gene mosaics, possessing material from *S. mitis* and *S. oralis* in pneumococci implies that these species are particularly important for inter-species recombination by *S. pneumoniae* (Humbert et al., 1995, Whatmore et al., 2000, Reinert, 2009). These species are thought to have played an important role in the acquisition of antibiotic resistance among pneumococci. Sibold and colleagues (1994) found that although the *S. oralis pbp2x* gene exhibited 20% genetic divergence relative to this gene carried in a penicillin susceptible R6 *S. pneumoniae*, mosaic *pbp2x* genes among clinically penicillin and cefotaxime resistant isolates showed only 3% divergence (Sibold et al., 1994).

Despite the implicated importance of *S. mitis* and *S. oralis* in the development of antibiotic resistance among pneumococci, the mode by which these species acquired resistance remains unknown, although is likely to have been facilitated by their carriage lifestyle in the oral cavities.

1.7 Genomics and Sequencing

1.7.1 First Generation Sequencing

Sanger and Coulson's (1975) plus and minus method for DNA sequencing marked the beginning of a revolution in genomics, and biology as a whole. The first step in this method involved the asynchronous polymerization of single stranded template DNA. This resulted in a mix of reads of different lengths, which were divided into 4 "plus" and 4 "minus" assays. In each minus assay, 3 of the 4 deoxynucleotide triphosphate (dNTP) nucleotides were included for polymerisation, whereas in the "plus" assays, the dNTP not present in the "minus" assay was used alone. Extension of reads in the minus assays consequently stopped at the site where the excluded dNTP should have been included, whilst the plus assays terminated unless the single dNTP could be added (Sanger and Coulson, 1975). A key innovation of this method was the use of an acrylamide gel to separate the products of the plus and minus polymerization reactions according to length, allowing the original sequence to be determined. Two years later Maxam and Gilbert published their competing sequencing method, which utilized site specific chemical cleavage of the DNA backbone, and subsequent fractionation of reads on a polyacrylamide gel for sequence determination (Maxam and Gilbert, 1977). Gilbert and Sanger would both share a Nobel Prize in Chemistry for their work on nucleic acid sequencing in 1980.

The Maxam-Gilbert method was initially popular, as it could be run using purified DNA- the Sanger method requiring reads to first be cloned to produce ssDNA. However, in 1977 Sanger and colleagues published a new method of sequencing, similar to the plus-minus method, but utilising base specific chain terminators. This method reduced the need to handle toxic chemicals and radioisotopes, and consequently, Sanger and colleagues' chain termination technique rapidly replaced the Maxam-Gilbert method (Schuster, 2008). Read length during the early development of the Sanger method rarely exceeded 25bp in length, and

only began to exceed 80bp with the development of dideoxy terminators (Schuster, 2008). The "Sanger sequencing" method would however dominate the field for much of the following 30 years.

The human genome project was a major driver in the further development of sequencing technology. The first automated sequencers became available in the mid-1980s, incorporating capillary electrophoresis into their design. Whole genome sequencing at this time typically involved the cloning of fragments of DNA into a bacterial vector for amplification. Individual templates needed to be purified, before fragments were then sequenced using fluorescent chain termination nucleotide analogues (Prober et al., 1987). Slab gel or capillary electrophoresis was then used to determine the sequence. This was the predominant method used to produce the first complete human genome sequence, the first draft of which was made available in 2000. Such techniques however came at a considerable expense. The National Human Genome Research Institute estimated the cost of the Human Genome Sequencing Project to be between US$10 and US$25 million (Margulies et al., 2005).

1.7.2 Second Generation Sequencing (Sequencing by synthesis)

The Roche companies 454 sequencer was the first of the second-generation sequencers to become available in response to the increasing demand for higher throughput technologies. Breakthroughs in emulsion based methods of DNA strand isolation, and the development of pyro-sequencing technologies at the picolitre scale were essential for this technology. Shendure and colleagues' (2005) "polony" method of multiplexing polymerase chain reactions (PCR) would soon lead to the development of a competing, high throughput method. As a result of these developments, multiples of 100,000 DNA templates could be sequenced simultaneously on picotitre plates or agarose thin layers. Such microarrays overcame a major limitation of the Sanger capillary

sequencing technology, which was limited to a maximum of 96 capillaries per machine (Schuster et al., 2007). As results of these breakthroughs, sequencing costs per base were rapidly cut, and a 100-fold increase in throughput was achieved (Margulies et al., 2005). This technology was also the first to remove the need for bacterial vectors during library preparation, simplifying the process (Margulies et al., 2005).

Reads were however shorter, at typically 100bp compared to the ~750bp reads produced by the later Sanger sequencing machines (Schuster et al., 2007), and they also had a greater per individual read error rate (Margulies et al., 2005). Those behind the Roche 454 sequencer argued that these shortcomings were compensated for by greater read depth. Despite these reassurances, and although an individual 454 sequencer was capable of achieving a throughput equivalent to 50 of the Applied Biosystem's 3730XL capillary sequencers, at approximately 1/6th of the cost, uptake remained slow. Concerns about the short read lengths, read fidelity, infrastructure costs and the ability to process the high volumes of data generated put off investors (Schuster et al., 2007). The technology was able to develop rapidly in response to these concerns however, and within 6 months of being on the market, read lengths had increased up to 250bp (Schuster et al., 2007). As a consequence, the technology began to become more widely adopted, and in recent years whole-genome sequencing has become a widely accessible technology across the globe.

1.7.3 Sequencing Overview

All second generation sequencing technologies follow a similar protocol: template preparation, sequencing, imaging (during which the nucleotide sequence is determined), genome alignment, and assembly.

Template preparation involves the random fragmentation of genomic DNA. The sequencing libraries generated subsequently come in two types, fragment templates (or "single-end reads"), and mate-pairs (or

"paired-end reads"). Fragment templates result from shearing genomic DNA into small sizes, typically <1000bp, before adapter oligos are attached to either end. Mate-pair/paired-end libraries result from selecting sheared genomic DNA sequence of a particular size, such as 2000bp, which are then circularised by attaching adapters to either end of the sequence. Fragmentation of these sequences, away from the joined adapters, subsequently yields linear mate-pairs or paired-end sequences, with adaptors then being ligated to the newly exposed ends (Murphy et al., 2012). These mate-pair/paired-end reads are often favoured as they offer additional distance and orientation information that can be used subsequently to aid *in silico* sequence assemblies. However, this method requires a higher amount of starting material, and consequently single-end reads may be favoured when starting material is limited (Wetzel et al., 2011).

Following library generation, NGS technologies typically involve the attachment or immobilization of spatially separated template DNA to a solid surface or support. Limitations in imaging resolution mean that sequences must undergo several rounds of PCR prior to sequencing, resulting in spatially separated, dense clusters of homologous DNA anchored to the support surface. In this way thousands to billions of sequencing reactions can be undertaken simultaneously (Metzker, 2010).

1.7.4 Roche/454

During library preparation for the 454 platform, A and B adaptors are ligated to the start and end of the target sequences. The adaptors contain universal primer sites (Metzker, 2010). The DNA is then split into ssDNA, before beads, coated in one of the complementary adaptors are introduced into the solution. Under conditions that favour the attachment of a single DNA strand per bead, DNA is then captured by the beads (Metzker, 2010). Emulsion PCR (emPCR) is used to isolate individual beads in a water-in-oil emulsion (Williams et al., 2006). Amplification and

enrichment of the beads by emPCR leads to a dense cluster of homologous DNA forming on the surface of the beads. The beads are then immobilised, either on a polyacrylamide gel, on a standard microscope slide (Polonator)(Shendure et al., 2005), through chemical crosslinking to amino-coated glass (Life/APG/Polonator)(Kim et al., 2007), or by falling into individual PicoTiterPlate (PTP) wells (Roche/454). The forward strands generated during emPCR are then removed using alkali denaturation (Leamon et al., 2003).

Each of 4 dNTPs is added sequentially to the immobilised beads, which release pyrophosphate when incorporated during DNA synthesis. The magnitude of the light is then used to indicate the number of bases being incorporated (Ronaghi et al., 1996). Sequencing then resumes until the next complementary base is added. The order and magnitude of the fluorescences are recorded in the peaks of flowgrams, calculated for each template cluster. The flowgrams can then be interpreted in order to determine the original template sequence. Current platforms generate relatively long reads (330 bases), improving the mapping of repetitive regions, and generating approximately 0.45Gb of data per run (Metzker, 2010). However, high error rates occur in homopolymer repeats due to the difficulties in detecting variations in fluorescence, and reagent costs are high. Consequently this technology is usually favoured when *de novo* genome assembly is required (Metzker, 2010).

1.7.5 Illumina (previously Solexa)

The current alternative to the 454 method, and most commonly used sequencing technology is the Illumina sequencer. This method is based on the generation of randomly distributed clusters of clonally amplified DNA fragments, which are attached to the surface of a glass slide; termed solid-phase amplification. The process begins by adding additional primers to the 3' and 5' ends of the template-DNA, the primer concentration regulating the number of clusters to be amplified. The

template-DNA is then washed over a glass slide, to which high densities of complementary forward and reverse primers are covalently attached, so that the template-DNA becomes anchored to the slides surface. Once attached, a polymerase makes a complement to this strand. The dsDNA is now denatured, and the original template strand washed away. The free end of the remaining linear DNA then loops to bind another primer on the plates surface, priming the PCR reaction. Following amplification, the newly synthesised dsDNA splits, creating linear ssDNA, which remains anchored to the slides surface. In this way growing clusters of linear ssDNA are generated, which are used to prime subsequent rounds of amplification via "bridge PCR" (Figure 13)(Metzker, 2010). An estimated 100-200 million spatially separated clusters can be amplified using this method of solid phase PCR (Illumina/Solexa).

Following PCR amplification, sequencing is undertaken using the method of cyclic reversible termination. This describes the cyclic addition of modified dideoxynucleotide triphosphate (ddNTP), which carry a hydroxyl group at the 3' position of the deoxyribose sugar, rather than a hydrogen atom. As a result, the ddNTP terminates DNA polymerase action until the nucleotide is removed by phosphorolysis, mediated by high ATP or pyrophosphate concentrations (Bentley et al., 2008). Under this method, a single ddNTP, carrying a nucleotide specific fluorescent molecule, is incorporated by the DNA polymerase as it moves along the primed template, preventing continued sequencing. The unincorporated ddNTPs are then washed away, before imaging occurs. The wavelength emitted by the fluorescent dye is then used to determines the base that has been incorporated. Cleavage of the fluorescent dye and reduction of the hydroxyl group then occurs, and following additional washing, sequencing is resumed (Metzker, 2010). Between 18Gb (single end) and 35Gb (paired end) of data can be generated by the Illumina Genome Analyser II platform, although read lengths are comparatively short, typically between 75 and 100bp in length (Metzker, 2010). As such the

Illumina technology has tended to be used when a reference sequence is already available, and *de novo* assembly is not required.

1.7.6 MiSeq

A major focus in the sequencing industry to date has been the production of platforms with increasingly larger throughputs, largely in response to studies such as the 1000 genomes project, which is aimed at better cataloguing genetic variation in humans. The reagent costs, throughputs, and support costs do not match the needs of small-scale projects, interested in sequencing smaller numbers of genomes. MiSeq is one of a number of recent technologies aimed at addressing this problem. Based on the Illumina HiSeq model of sequencing by synthesis (Bentley et al., 2008), this sequencer reduces run times from 1.5 days to 4 hours for 36-cycle sequencing, or from 8.5 days to 16 hours for 200-cycle sequencing, due to improvements in reducing imaging time, faster microfluidics and reduced size of the flow cell (Loman et al., 2012). Furthermore, whilst HiSeq and MiSeq are capable of generating reads up to 150bp long, longer reads are better suited to the MiSeq setup (Quail et al., 2012). Additionally the MiSeq shows a low read error rate (indels/mismatches) and more accurate SNP calling, compared to other available sequencers (Quail et al., 2012).

Sequence yields are substantially lower 1.5-2Gb per run (Quail et al., 2012), costed at $502 per Gb, compared to 600Gb produced for $41 by the latest Illumina HiSeq 2000. However, despite these limitations, this technology remains competitive for small sequencing projects.

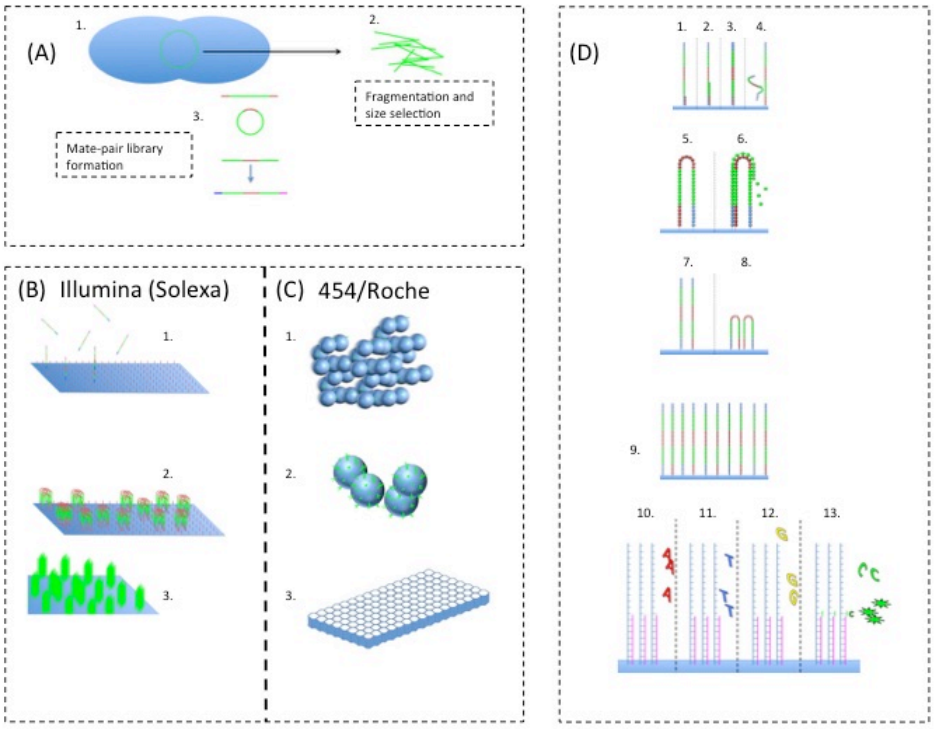


Figure 13: Sequencing technologies compared. A: The bacterial genome is extracted (1), fragmented and fragments are size selected for sequencing (2). A mate pair library may be generated (3). In this case adapters are annealed to both ends of the DNA fragment. The fragment is circularised, before a second round of fragmentation, away from the adaptor sequences, results once more in a linear sequence, with an adaptor sequence in the central region. A second pair of adaptors are then attached to each of the fragment ends, specific to the sequencing platform being used. B: During the Illumina (Solexa) process DNA fragments are washed over a plate coated in complementary adaptors (1), to which the fragments anneal. PCR leads to dense, spatially separated clusters of identical DNA fragments (2), which are then sequenced (3). C: During the Roche/454 method, DNA fragments are attached to beads coated in the complementary adaptor, under conditions favouring the binding of one fragment per bead (1). PCR amplification occurs (2), before beads are washed over a microtitre plate, with each bead falling into a well for sequencing. D: Illumina sequencing in detail. Following attachment of linear DNA to the microtitre plate, PCR of the attached DNA, and removal of the unbound strand, leads to a single strand of anchored DNA ready for sequencing (1-4). Bridge-sequencing then follows, during which the anchored DNA folds to attach to an additional oligo (5), before it is amplified (6), and the strands separate (7). This process is then repeated several times (5-8), before sequencing occurs (10-13).

1.7.7 Sequencing conclusion (Third generation)

Over 30 years since it's development , the Sanger method remains in use today- the read lengths still not matched by any of the second generation sequencing technologies (SGST). It also requires less starting material, does not use PCR, during which mutational errors can arise, and is less susceptible to GC-rich, AT-rich biases and subsequent errors. In addition, a number of error mechanisms, unique to SGSTs, have been found to occur. One such case is that of dephasing, which occurs during amplification of clonal clusters. Here, when there is incomplete extension

of the lagging strand, lagging strand dephasing occurs, or in the reverse case, the addition of multiple bases during a single replication cycle causes leading-strand dephasation. Irrespective of the mechanism, dephasation increases the fluorescent noise, resulting in an increase in base-calling errors, and a reduction in read lengths (Erlich et al., 2008).

The errors associated the SGST have increased the demand for a third generation of sequencing technologies. Many of the aims of these technologies are based on producing longer read lengths, and utilising less starting material, but generating read data at the same speeds associated with SGST technologies (Schadt et al., 2010).

Nanopore sequencing is one method that may populate the third generation of sequencers. Nanopore technology offers the potential for single molecule sequencing, a highly desirable feature of future sequencers for clinical diagnostic settings. Here DNA is unzipped, and the ssDNA strand fed through nanopore. By passing an ionic charge through the nanopore, alterations in current, caused by nucleotides or amino acids as they are passed through the pore, allow the molecule to be identified (Schneider and Dekker, 2012).

1.8 Summary

The breakthroughs in sequencing technology discussed above have allowed researchers to study bacteria at the population level. The resulting and continued explosion of sequence data has given novel insight into the mechanisms by which such populations evolve, and in particular, respond to clinical pressures.

Such studies have allowed a greater understanding and characterisation of the role that recombination has had on the population biology of the pneumococcus (Croucher et al., 2011). It is now possible to assess the mechanism of beta-lactam resistance genome-wide, rather than being

restricted to particular loci (Chewapreecha et al., 2014a). In addition sequencing has enhanced surveillance, allowing epidemiologists to better understand the lineages present, and the relative risks that these may pose to communities, guiding clinical interventions (Everett et al., 2012).

Despite the broad application of this technology, the mechanism of beta-lactam resistance prevalent among globally important MDR lineages such as PMEN1 remain unknown. There is similarly a deficit in knowledge as to the role that *S. mitis* and *S. oralis* have played in the development of resistance among pneumococci, including the frequency with which such species are able to exchange genetic material. The wider role that recombination plays in repairing genetic damage, and the effect of oxidative stress on genomic integrity of the pneumococcus similarly remains poorly investigated using these technologies.

The analyses discussed herein take a next-generation sequencing approach to understanding the role that recombination has had in genomic repair, and the development of beta-lactam resistance in the pneumococcus.

1.8.1 Project Aims

The aim of the project was to:

- Better determine the mechanisms by which beta lactam resistance emerges clinically
- To characterise the occurrence of genetic damage, and the role of recombination in deletion repair and dissemination of beta lactam resistance

2 Materials and Methods

2.1 Overview

This study was a collaborative project undertaken between the University of Liverpool, Liverpool, the University of Warwick (UoW), Coventry, and the Wellcome Trust Sanger Institute (WTSI), Hinxton Cambridge. The projects described herein availed of sample repositories based onsite at the WTSI, UoW and the Malawi-Liverpool Wellcome (MLW) Clinical Research Programme (www.mlw.medcol.mw), located in Blantyre, Malawi. MLW provides routine diagnostics services to patients admitted to QECH, Malawi's largest hospital. QECH, a publicly funded hospital, receives approximately 50,000 patients annually, and is estimated to serve a population of over 1 million (Everett et al., 2011). The MLW repository currently holds over 5000 pneumococcal samples, which have been collected since 1996.

Laboratory work was undertaken across three sites. The work undertaken at MLW availed of the diagnostics facilities and large sample repository onsite. At UoW medium-throughput MIC testing, fluorometry and DNA extraction was undertaken. Fluorometry and DNA extraction was also undertaken at the WTSI. Bioinformatic support and training was provided both remotely and while on site by WTSI, and the Quantitative Biology Centre (QuBIC) group based at UoW.

2.2 Clinical samples (MLW)

2.2.1 Blood culture

Blood samples obtained from Malawi were collected through venepuncture by medical staff at QECH in accordance with the MLW standard operating procedures MLW.SOP.C.8.000 Version 1 and MLW.SOP.C.001 Version 2 (Appendix 10.1). Briefly, patients were identified for blood culture in cases of fever (>37.5^{o}C) or abnormally low body temperature (<35^{o}C), and where a clinical suspicion of bacterial infection was identified, such as lacking a positive smear test for malaria. Blood was also taken from patients who remained febrile

following initial malaria treatment. In cases where samples were taken, attempts were made to ensure that sampling was taken prior to antibiotic treatment.

Blood sampling obtained 5-10mL of blood from adult patients (15yrs≤), which was collected in a blood BacT/Alert SA blood culture bottle (AB BioMérieux, Solna, Sweden). In children (<15yrs) 1mL of blood was taken and stored in BacT/Alert PA blood culture bottles (AB BioMérieux). Once blood was drawn it was sent to the MLW diagnostics laboratory for processing (Figure 14).

On receipt, the blood culture bottle was screened for bacterial growth using an automated BacT/Alert 3D (AB BioMérieux) system. Negative cultures were discarded after 7 days (adults) or 5 days (children), whilst positive cultures were sent for further diagnostic testing (Figure 14).

2.2.2 Cerebrospinal fluid

Cerebrospinal fluid (CSF) samples were collected though lumbar punctures (LP) by QECH staff using MLW.SOP.C.8.000 Version 1 (Appendix 10.2). CSF samples were taken from children following blood sampling (section 2.2.1), and adults with suspected meningitis. At least 2mL (ideally 5mL) of CSF was collected in a universal container (Thermo Fisher Scientific, MA, USA) before being sent for processing.

Samples were processed by centrifugation at 12,000rpm for 10 minutes, which separated any suspension in the CSF sample. The supernatant was then discarded in 1% Virkon solution (DuPont, Delaware, USA), and the pellet sent for further testing (Figure 14).

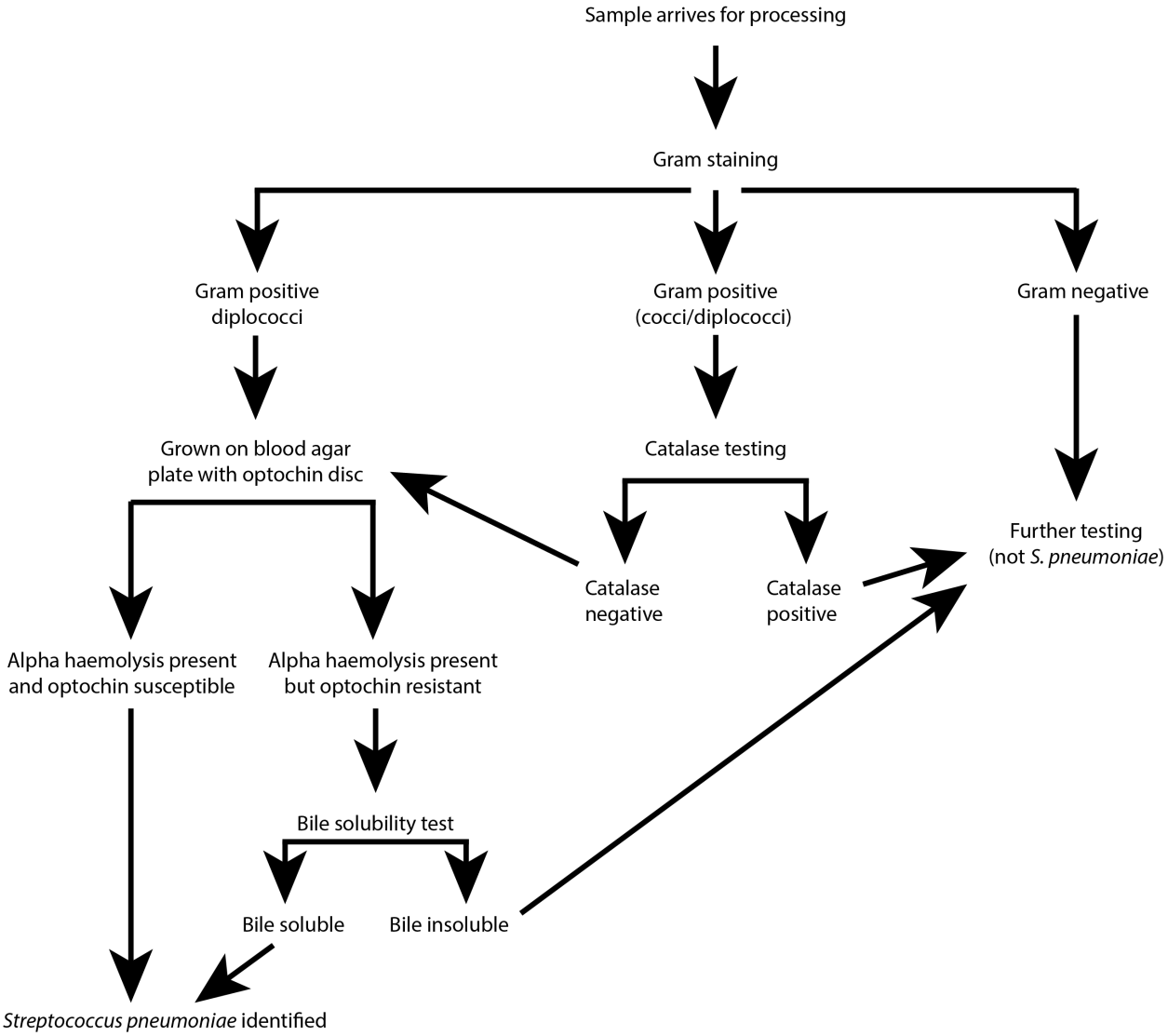


Figure 14: Processing of samples collected at MLW by the diagnostics team.

2.2.3 Nasopharyngeal swabbing

Nasopharyngeal samples were collected by QECH staff between 2001 and 2010 as part of a vaccine efficacy trial conducted in Malawi (French et al., 2010). Adults (≥15yrs) were recruited into this study based on having recovered from an IPD event, with a random subset of these subsequently receiving PCV7.

Nasopharnygeal samples were taken from such patients by passing a per-nasal swab through the nostril whilst exerting an upward pressure on the end of the nose, which facilitated access to the anterior nose. Once the swab had reached the posterior wall of the nasopharynx, approximately half the distance between the nostril and the earlobe, it was left in place for up to 5 seconds before being

withdrawn and placed in STGG media for storage at -80°C (Figure 15)(Appendix 10.3)(Gibson and Khoury, 1986, O'Brien et al., 2003).

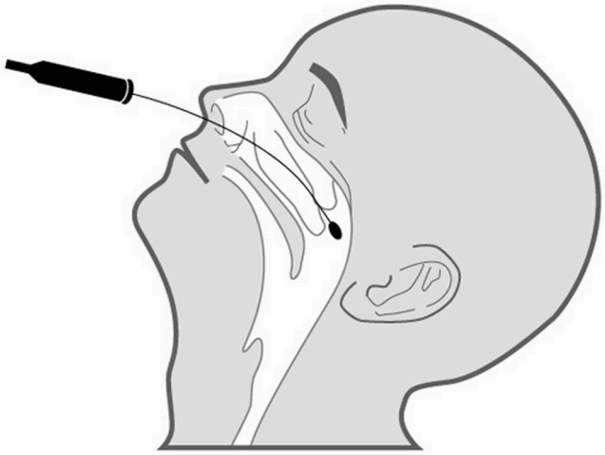


Figure 15: Depiction of the per-nasal swabbing procedure, in order to sample bacteria growing at the back of the nasopharynx.

2.2.4 Nasopharnygeal aspirates

Nasal aspirates were collected by MLW staff between late-2010 and 2012, using an Argyle™ suction catheter (Medtronic, Minneapolis, USA)(Appendix 10.4). The patient's head was tilted into a slightly hyper-extended position, and the catheter inserted into the patients left nostril until it reached the posterior of the nasopharynx. Suction was applied at 13kPa (100mmHg) before the catheter was removed. The 3mL sample was then deposited in a 5mL plastic vial, which was delivered to the diagnostics laboratory.

On receipt of the aspirate, the sample was divided equally between two cryovials, one containing universal transport medium (UTM), and the other, skim-milk, tryptone, glucose, and glycerine (STGG)(Appendix 10.5) containing cryovial. The STGG sample was incubated at 37°C, 5% CO_2 overnight, before being stored at -80°C. UTM samples were used for influenza virus (H1N1) screening before also being stored at -80°C.

2.2.5 Sample collections

PMEN (section 1.3.8) lineages were studied extensively throughout, and included four separate datasets: PMEN1 (Croucher et al., 2011), PMEN2 (Croucher et al., 2014b), PMEN14 (Croucher et al., 2014a) (Table 4), and PMEN27 (Everett et al., 2012).

The PMEN1 (Spain^{23F}-1) collection consisted of 240 whole genome sequences (WGS) isolated between 1984 and 2008. From this collection 175 PMEN1 samples were available for laboratory study. In addition, a pre-PMEN1 isolate (sequence and sample), BM4200 (4232_7#1), isolated from France in 1978 was included for comparative purposes. The PMEN1 collection represented a global collection of isolates, confirmed by MLST to be members of the PMEN1 clone (ST81, commonly serotype 23F).

PMEN2 (Spain^{6B}-2) is characterised by being of sequence type (ST) 90 and typically carrying a serotype 6B, having arisen in Spain in the 1960's (McGee et al., 2001, Croucher et al., 2014b). The samples used in this study included 80 WGSs for strains isolated between 1988 and 2009. Isolates were confirmed by MLST to be part of the PMEN2 lineage, and were collected globally, although greater sampling was undertaken in Iceland (45%) where this lineage was particularly prevalent during the 1990's (Croucher et al., 2014b). Phage insertion was documented across this lineage similar to PMEN1. Of particular note, was the insertion of a prophage, designated ΦIC1, into the *comYC* (also known as *comGC*) gene. Disruption of the *comYC* gene, which encodes a major pilin protein, has the potential to severely disrupt recombination, by preventing construction of the pilus structure necessary to interact and possibly draw-in DNA necessary for transformation (Croucher et al., 2014b, Muschiol et al., 2015).

PMEN14 (Taiwan^{19F}-14) was originally characterised as belonging to ST236, and typically carrying a serotype 19F capsule. The isolates included herein were identified as belonging to ST236, and included 35 globally sampled isolates and 87 isolates from the Maela refugee camp, located on the Thailand-Myanmar border (Turner et al., 2012). An additional four different ST (non-PMEN) datasets collected from the Maela camp were also included. Here, a dense

sampling approach was taken, during which pneumococci were isolated from cases of asymptomatic carriage. Collection was undertaken on a cohort of approximately 1000 infants and a quarter of their mothers over 3 years, between 2007 and 2010. From this repository 3,085 pneumococci were subjected to WGS at the WTSI- so that almost 100 isolates were sampled from each consecutive month of the sampling period. The clustering program Bayesian Analysis of Population Structure (BAPS) was applied to the dataset, allowing clusters of isolates sharing sequence similarity to be delimited. From these, four of the largest clusters, containing sequences determined to be of the same ST were selected for study in this analysis (Chewapreecha et al., 2014a).

All of the above PMEN lineages, and the isolates sampled from the Maela refugee camp, have been characterised in terms of its recombination, mutation, divergence patterns, and mobile elements previously (Croucher et al., 2011, Chewapreecha et al., 2014a).

A set of 60 serotype 1 WGSs were also included from Malawi. These belonged to ST217, typically serotype 1, which is a frequently associated with IPD in Malawi. This lineage has been identified as a member of the PMEN27 (Sweden^{1}-27), although this lineage is not associated with MDR (Everett et al., 2012). In addition a mixed set of 609 WGS pneumococci were similarly included from Malawi. Reference and annotations for the Malawian pneumococci datasets were provided based on previous analyses (Everett et al., 2012). For convenience, a summary of the sequence datasets used in the analyses herein is provided in Table 4.

Publicly available archives (www.ebi.ac.uk/ena) were used to obtain partial and WGSs for additional streptococcal species (chapters 4 and 5). For the laboratory assays conducted into the effect of oxidative stress (chapter 6) clinical sample collections of UK isolates, based at the UoW, were also used.

Dataset	Number of samples
PMEN1	240
PMEN2	80
PMEN14	122
14 ST63	66
19F ST4414	245
23F ST802	127
23F ST4413	83
19A/15B/C ST199	94
1 ST217	60
Malawi (total)	669
0100993	1
37A	1

Table 4: A summary of the pneumococcal sequence datasets used in the following projects. Numbers of genome sequences available in each dataset is also indicated.

2.3 Laboratory protocols

2.3.1 Media

Brain Heart Infusion (BHI) agar and broth (Oxoid, Basingstoke, UK) were used for the standard culturing of pneumococci. BHI contains a mix of brain infusion and brain heart infusion solids, which combined with the additional constituents achieves a buffered medium of approximately pH 7.4 (Thermo Scientific, 2015). As BHI contains organic components, the exact chemical constituents are variable between batches, and it is therefore defined as a complex medium.

BHI agar was enriched for pneumococcal growth using 5% defibrinated horse (Thermo Fisher Scientific) or sheep blood (provided onsite by MLW), which can be used interchangeably as a source of catalase (CDC, 2011). Catalase is important for the breakdown of hydrogen peroxide (H_2O_2), which can otherwise lead to cell lysis and inhibited growth during aerobic growth (Pericone et al., 2000). Iso-sensitest (IS) agar and broth (Oxoid) was used for antimicrobial susceptibility testing according to published guidelines (BSAC, 2014), and was similarly enriched with blood during agar preparation.

Mitis Salivarius Agar (Neogen Corporation, Ayr, UK) was used for the isolation of mitis group bacteria from mixed samples. Mitis Salivarius Agar was

supplemented with 1% tellurite (Thermo Fisher Scientific) according to the manufactures instructions.

All isolates were grown under aerobic conditions at 37°C and 5% CO_2 for 18-20 hours unless stated otherwise. Conditions for microaerophilic growth were achieved by culturing isolates in a candle jar with a CampyGen Compact sachet (Oxoid).

2.3.2 Storage

Isolates were stored in 2mL Cryogenic Vials (Sigma-Aldrich Inc, Dorset, UK) containing 50% glycerol-BHI, or using MicroBank™ bacterial preservers (Pro-Lab Diagnostics, Neston, UK). All samples were stored at -80°C.

2.3.3 Microbial identification tests

Optochin testing was performed for the presumptive identification of pneumococci and to check for contamination (Figure 16a for positive identification of a pneumococcus). Optochin testing (section 1.3.1) is commonly used to identify and isolate pneumococci under the general principal that this species is optochin sensitive, whereas closely related mitis group streptococci (e.g. *S. mitis,* and *S. oralis*) are optochin insensitive. Briefly, a 5µg/mL optochin impregnated disc (Oxoid) was placed in the initial streak of sample on a blood agar plate, and incubated overnight under standard conditions prior to examination. The presence of alpha-haemolytic colonies, and a zone of inhibition ≥14mm in diameter around the optochin disk (Oxoid) following growth were interpreted as indicative of pneumococcal growth (CDC, 2012).

In addition, pneumococci characteristically form small moist or mucoidal colonies (Allegrucci and Sauer, 2007)(Figure 16), similar to other alpha-haemolytic commensal streptococci (viridans group streptococci). Following culture of between 24 and 48 hours pneumococcal colonies however appear flattened, with a depression of the central portion of the colony, which does not occur among viridans streptococci. Colony morphology and optochin sensitivity

were consequently used during identification of mitis group isolates from blood agar plates (CDC, 2012).

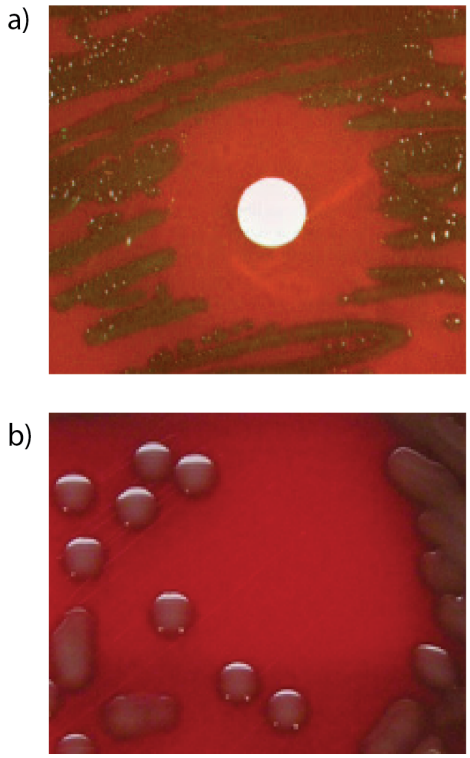


Figure 16: Figures indicate (a) the classic zone of optochin inhibition characteristically used for the putative identification of pneumococci, and (b), the large "mucoidal" colonies often formed by this species (image sources: (a) http://bmtjournal.blogspot.co.uk/2007/12/particulars-of-patient-name-tong-wei.html, and (b) http://imagestack.co/36951262-streptococcus-pneumoniae-optochin-test.html).

2.3.4 Gram stain

A sterile swab (TexWipe®, Kernersville, USA) was used to transfer colonies from the surface of a blood agar plate culture, to create a thin smear on a clear glass microscope slide (Thermo Fisher Scientific), approximately 1cm^2. The cells were allowed to air dry on the surface of the slide, and passed through a Bunsen flame- this denatures surface proteins, helping the cells to stick to the surface of the glass slide, preventing their being washed-off subsequently (heat-fixation).

A Pasteur pipette (Thermo Fisher Scientific) was used to flood the slide with crystal violet solution, and left for 1 minute to penetrate the cell walls. Excess crystal violet is rinsed off with water, and iodine solution poured onto the slide

and left for a minute, forming crystal violet-iodine complexes inside the bacterial cells. Decolouriser was poured over the slide for 1 to 5 seconds, which easily penetrates the thin peptidoglycan walls (section 1.5.1) of gram-negative bacteria, removing the crystal violet-iodine complexes. The decolouriser was then washed off with water. The slide was flooded with safranin for 30 seconds, re-staining the gram-negative cells. Excess safranin was washed off using water and the slide allowed to dry before being viewed under a microscope at high (oil emersion) or low magnification. Colonies stained violet were recorded as gram-positive, whilst red/pink colonies indicated gram-negative organisms present (Figure 17).

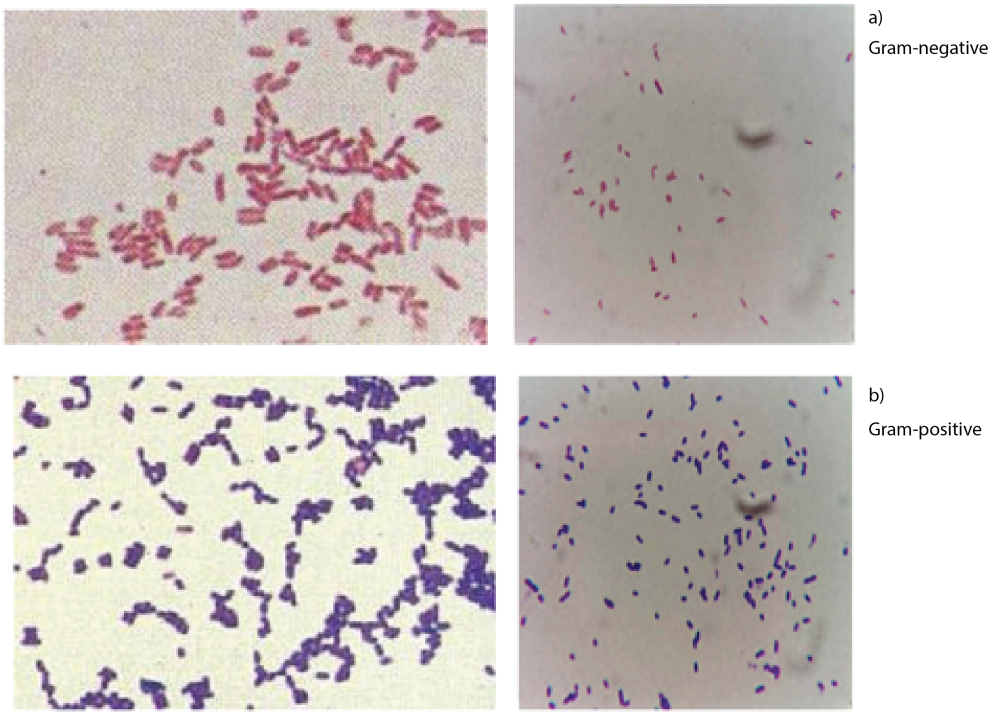


Figure 17: Gram negative (a) and gram positive (b) colonies viewed under microscope (source: http://biology.clc.uc.edu/fankhauser/labs/microbiology/gram_stain/gram_stain.htm).

2.3.5 Catalase testing

A sterile swab (TexWipe®) was used to transfer colonies onto a microscope slide (Thermo Fisher Scientific). Approximately 1mL of 3% hydrogen peroxide (Sigma Aldrich Inc.) was added to cover the smear. Catalase positive bacteria breakdown the hydrogen peroxide, yielding water and gaseous oxygen, which is

observed as bubbles, whereas for catalase negative bacteria such as *S. pneumoniae* no gas is produced (Figure 18)(CDC, 2012).

Figure 18: Results from catalase testing. In the right-hand figure, catalase breaks down hydrogen peroxide yielding oxygen, resulting in bubbling. In the left-hand figure no catalase is present, and therefore no bubbling is seen (source: http://microbeonline.com/catalase-test-principle-uses-procedure-results/).

2.3.6 Spectrophotometer readings

Spectrophotometer readings were taken using an iEMS™ platereader (MTX Lab Systems Inc, Vienna, USA) at the UoW, and a FLUOstar® Omega (BMG Labtech, Ortenberg, Germany) at the WTSI. Readings were taken at OD_{600} every 15 minutes unless stated otherwise, with a 10 second shake prior to a reading being recorded. Plates were incubated for 16 to 24 hours at 37°C, in 96 well plates (Thermo Fisher Scientific), which were sealed using adhesive film (Thermo Fisher Scientific). Condensation, which can invalidate OD readings, was avoided as both spectrophotometers used a heated-lid system, preventing moisture condensing on the seal of the 96 well plates.

2.3.7 Antimicrobial susceptibility testing

Antimicrobial susceptibility testing protocols and interpretive guidelines are published by three main institutions: the British Society for Antimicrobial Chemotherapy (BSAC) (BSAC, 2015), the Clinical Laboratory Standards Institute (CLSI) (CLSI, 2015) and the European Committee for Antimicrobial Chemotherapy (EUCAST) (EUCAST, 2015). BSAC guidelines and breakpoints were followed, unless indicated otherwise. MIC values were recorded to indicate

antibiotic susceptibility, which defines the minimum concentration of antibiotic required to inhibit bacterial growth (Jorgensen and Ferraro, 2009).

2.3.8 Plate preparation (for MIC Etest®)

IS agar (Oxoid) was prepared according to the manufacturers guidelines and supplemented with 5% sheep or defibrinated horse (Thermo Fisher Scientific) or sheep blood for pneumococcal growth. Prepared media was poured onto sterile petri plates to a depth of approximately 4mm (equivalent to 25mL per 90mm petri dish). Plates were prepared in a safety cabinet to avoid contamination, and dried to remove excess moisture. Plates were stored at fridge temperature (4-8^{o}C) for up to a week prior to use (BSAC, 2013). Control strains used during antibiotic susceptibility testing are indicated in Table 5, and were chosen in accordance with BSAC and CLSI guidelines.

Species	Control used
S. pneumoniae	NCTC 12977 (ATCC 49619)
S. pneumoniae	ATCC 700669
S. constellatus	NEQAS collection
S. mitis	NEQAS collection

Table 5: Control strains used for antibiotic susceptibility testing.

2.3.9 Inoculum preparation

Samples were incubated overnight, in the presence of an optochin disc (Oxoid) as described above. Following incubation several colonies were picked using a 10µL sterile loop (Thermo Fisher Scientific), and resuspended in phosphate buffered saline (PBS)(Appendix 10.6) to a concentration that matched or exceeded that of a 0.5 McFarland standard, equivalent to a cell count density of approximately $1.5x10^{8}$ (Pro-Lab Diagnostics, 2012). The resulting suspension was used within 15 minutes of preparation.

Plates were inoculated by spreading the prepared inoculum evenly over the surface of the agar plate, using a sterile cotton swab (TexWipe®). An additional drop of PBS could be used to ensure a more even spread of inoculum. The plate was left to dry for a couple of minutes, before an Etest® strip (AB BioMérieux,

Marcy l'Etoile, France)(Figure 20) was applied, face up, to the agar surface, using a pair of sterile tweezers. By running the forceps along the length of the strip, from low to high concentration, underlying air-bubbles were removed. Plates were then incubated overnight, in stacks approximately 6 high, to ensure sufficient temperature penetration (BSAC, 2013).

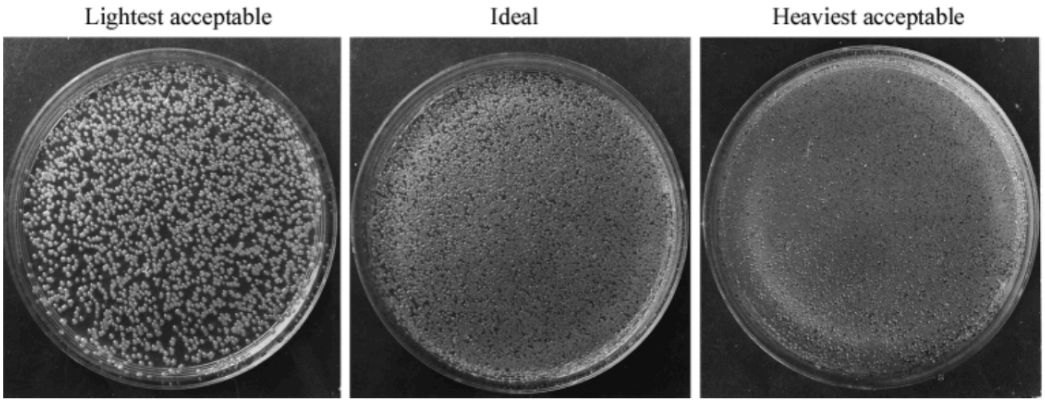


Figure 19 The inoculum that yields a semi-confluent growth of colonies should be chosen (Andrew et al., 2001). Denser inoculums yield reduced zones of inhibitions, while lighter inoculums lead to greater inhibitory zones.

Following incubation, plates were inspected for uniform growth (Figure 19) and MIC values recorded in accordance to AB BioMérieux guidelines:

- The Etest® is read using the point where growth touches the strip
- Haemolysis, and micro-colonies are ignored
- The upper MIC value should be taking when growth occurs between two markings
- Growth over the highest value, or below the lowest value is recorded as greater than or less than that respective value
- Water on the surface of the strip can cause a thin line of colonies to grow next to the strip, but this should be ignored

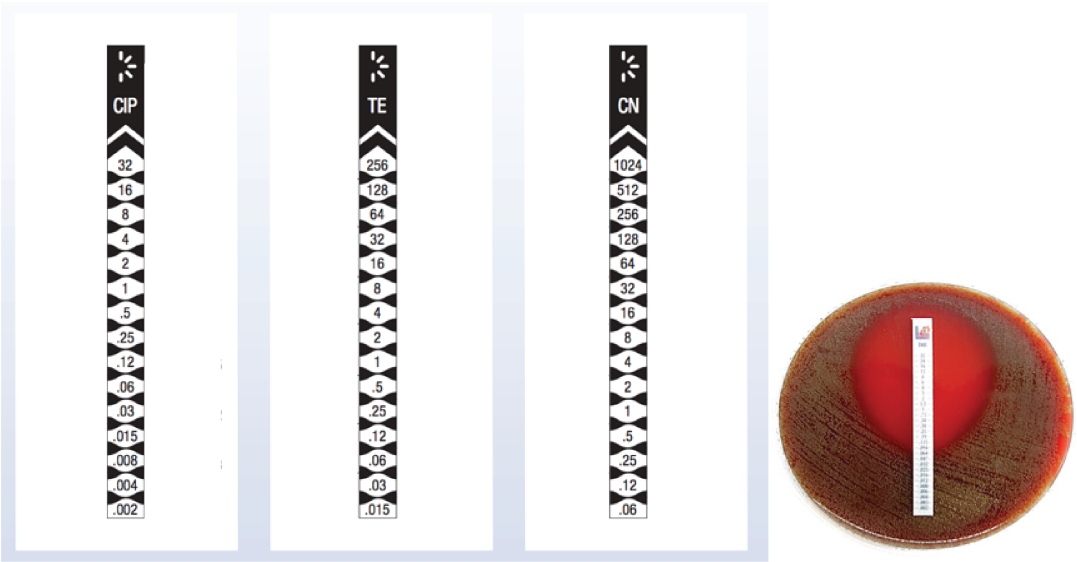


Figure 20: The appearance of the Etest® and correct placement on a blood agar plate (pictures taken from //www.oxoid.com/pdf/uk/MICE-reading-guide-A5-march08.pdf, and http://uni-chem.rs/images/news_mic_test_strip.jpg).

2.3.10 Plate dilution method

The plate dilution series was prepared following standard practice (Wiegand et al., 2008). Isolates were inoculated into 200µL of BHI broth in a 96 well plate, and sealed with film. Following incubation in a plate reader, absorbance measurements allowed cell densities to be standardised to approximately OD_{600} 0.8-1 (equivalent to a 0.5 McFarland standard), by the addition of BHI (Oxoid) or bacterial colonies from plate growth. Plates were then used immediately following preparation.

Antibiotics were sourced from Sigma-Aldrich Inc, and were prepared from powder according to the manufacturers instructions. Briefly, an appropriate potency of antibiotic was achieved by dissolving in distilled water, which was facilitated by gentle heating and stirring where necessary. The antibiotic was added to 5% blood and mixed by inversion, prior to being mixed to the IS agar (Oxoid) and poured into petri plates (Sigma-Aldrich Inc). Antibiotics were prepared on the day of use to avoid loss of activity. A microbiological replicator tool (Baryshnikova et al., 2010) was then used to inoculate 48 strains at a time onto an antibiotic containing blood-agar plate. These were incubated overnight before bacterial growth was assessed. Two control plates containing no

antibiotic were also inoculated to ensure sufficient growth of each isolate from the 96-well plate. Plates were assessed for growth following incubation, and an MIC recorded for each isolate.

2.3.11 Continuous growth apparatus

Continuous growth experiments (Figure 21) were undertaken in accordance with the procedure described by Waite and colleagues (2001, 2003).

A 10x20mm cylindrical sorbarod filter (Sigma Aldrich Inc) was inserted into the bottom of a plastic cylinder of equal diameter and a length of 500mm. The unit was sealed at the top with a rubber bung, through which a disposable hypodermic needle was driven. The needle was itself attached to a length of 1.5mm bore silicone tubing, allowing BHI media (Oxoid) to be supplied by action of a periplasmic pump, the other end of the tubing attached to a 500mL Schott bottle (Sigma-Aldrich Inc) containing BHI media (Oxoid). In this way, fresh media could be continuously supplied to the top surface of the sorbarod filter, and effluent from the bottom of the sorbarod filter (Sigma Aldrich Inc) could be collected in a replaceable 300mL glass bottle (Figure 21).

An additional hypodermic needle was driven through the side-wall of the 10x500mm plastic cylinder, through which the initial bacterial inoculum could be delivered to the top surface of the sorbarod filter (Sigma Aldrich Inc). The whole apparatus was sterilised by autoclaving at 121^{o}C for 15mins.

Bacterial stocks were prepared to a concentration of 1x10^{6} CFU using the Miles and Misra technique (Miles et al., 1938) prior to use in sorbarods, and 1mL of

this preparation used to inoculate the top surface of the sorbarod.

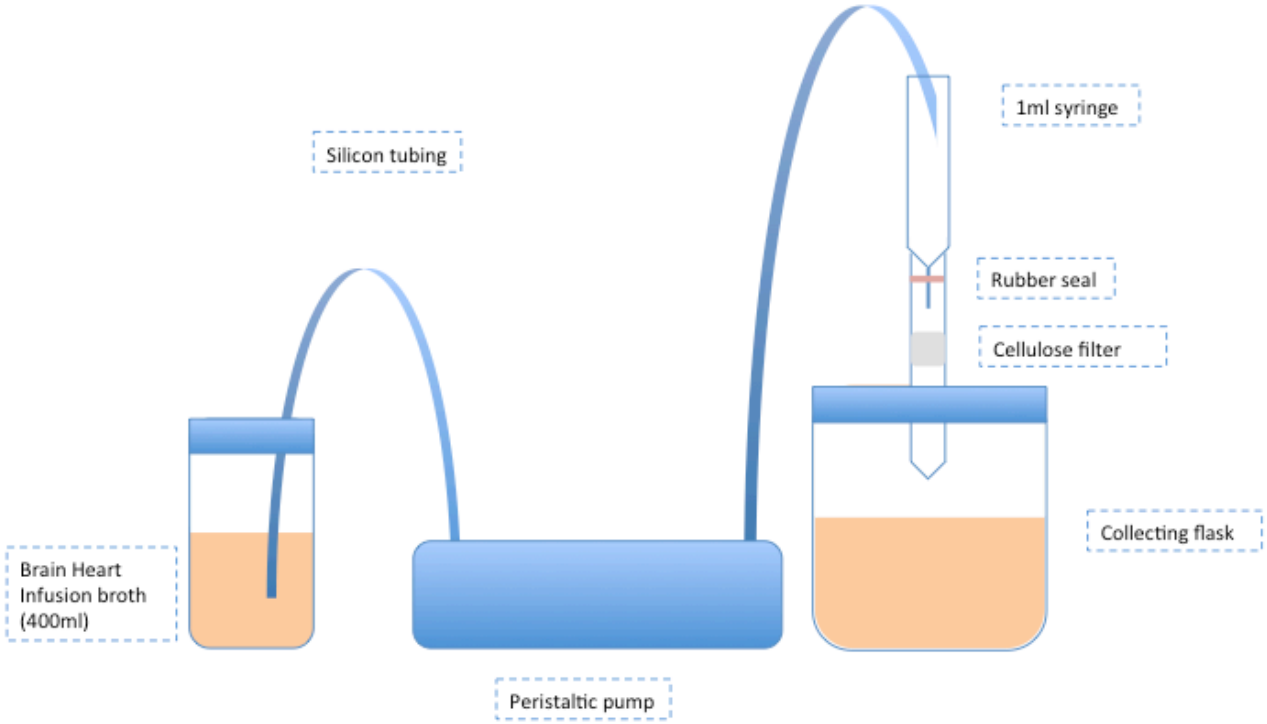


Figure 21: The assembled sorbarod apparatus. At the left-hand side the silicon tubing feeds nutrient broth, through the action of the peristaltic pump, onto the top surface of the sorbarod filter. Effluent media is then falls into the collecting flask as nutrient broth filters through the sorbarod.

2.3.12 Hydrogen peroxide MIC testing

MIC assays were carried out in a 96 well plate (Thermo Fisher Scientific), with a dilution series created of adding hydrogen peroxide to distilled water at varying concentrations. 20µl of each dilution was added to 180µl of BHI media (Oxoid), containing 10^4 cfu/mL of bacteria. The 96-well plate was then sealed, and incubated in a spectrophotometer for 16 hours, taking measurements at OD_{600} every 10 minutes.

2.3.13 Amplex® Red assay

The Amplex® Red reagents (Life Technologies, Warrington, UK) consists of 10mM of 10-acetyle-3,7-dihydroxyphenoxazine, and horseradish peroxide. 5µl of 10mM stock concentration of Amplex® Red dissolved in Dimethyl Sulphoxide (DMSO) was added to 5µl of 10U/mL stock of horseradish peroxide. To this, 970µl of PBS was added, and a further 20µl of sample, so that each well of the 96-well plate had a total volume of 200µl. Hydrogen peroxide present in the sample is converted by horseradish peroxide into water using Amplex® Red as an electron donor (Figure 22). Resorufin, resulting from this reaction is highly coloured, and can be recorded in a spectrometer. As such, absorbance was measured at 555nm on a UV spectrophotometer. Hydrogen peroxide production for isolates was calculated by cellular supernatant as a substrate for the peroxidase enzyme. Briefly, 500µl of media effluent was collected from the sorbarod apparatus. This was centrifuged and the supernatant added to the Amplex® Red reagent (Life Technologies). The Amplex® Red assay was then performed in a spectrophotometer at 34°C (being representative of the nasopharyngeal environment), at ~pH7.4.

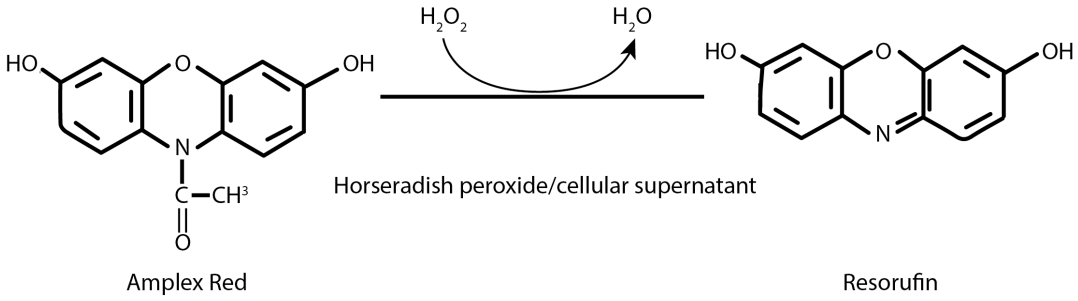


Figure 22: The reaction undertaken in the Amplex® Red assay. Using horseradish peroxide or cell supernatant, Amplex® Red acts as an electron donor for the reduction of hydrogen peroxide to water, producing fluorescent resorufin (image: http://www.biotek.com/assets/tech_resources/124/synhtfig1.gif).

2.3.14 Transformation experiments

S. pneumoniae strains were cultured on blood plates as previously described. Following overnight growth, a single pneumococcal colony was picked and used to inoculate pre-warmed 10mL BHI broths (Oxoid), which were then incubated

until an OD_{600} was achieved equivalent to the bacteria reaching log phase growth. This was found to be 0.3 for ATCC700669 and the serotype 3 strains used in this study. The OD_{600} could be measured periodically by pipetting 1mL of sample and measuring the OD_{600} on a Qubit® spectrometer (Life Technologies). Prior to the addition of CSP, 20µl of pre-transformation broth was streaked over a control plate. 1mL of the remaining broth was then used to inoculate a fresh 15mL tube (Thermo Fisher Scientific), and 2µl of CSP (5µg/mL)(Appendix 10.7) was added, alongside 5µl calcium chloride (500mM), and at least 200ng of the DNA that was to be used to transform the host cells. An extra 1mL of BHI broth (Oxoid) was added and this was incubated at 37°C for an hour (a "resting" period). 20µl of broth containing the transformed bacteria was then spread on an antibiotic plate. The transforming DNA typically carries an expressed phenotype, such as an antibiotic resistance cassette, which can then be used to determine whether the transformation has been successful or not.

2.3.15 Miles and Misra (1938)

Bacteria were scraped from the surface of an agar plate and resuspended in PBS to a concentration equivalent to a 0.5 McFarland standard (Pro-Lab Diagnostics, 2012). Diluting this suspension by the ratio 1:10 created a dilution series. At each dilution 10µl of sample was then spread evenly over the surface of an agar plate. By counting the number of colonies that grew on the plate following incubation, the number of colonies in the original sample could be determined. Colony counting was carried out as soon as separate colonies were visible and it was practical, in terms of colony density. To check for consistency, colony counting was undertaken at least twice, over two different dilutions. In this way, by assuming each colony resulted from a single bacterium, the number of bacteria in the original 10µl used to inoculate the plate could be calculated, and the corresponding dilution value was used to estimate the CFU present in the original sample.

2.4 PCR

2.4.1 Sample preparation

Samples were grown overnight in broth or plate culture and checked for purity. Up to three colonies were picked using a 10µl plastic loop (Thermo Fisher Scientific) and resuspended in 100µl of TE Buffer (10mM Tris-HCL, 1mM EDTA, pH 8.0) or deionised water.

For colony PCR, colonies were picked from the surface of an agar plate, resuspended in PBS, and a 100µl volume transferred to a sterile 1mL tube (Eppendorf, Hamburg, Germany). To achieve cell lysis, the suspension was then placed in a heat block at a temperature of 105^{o}C for 15 minutes. Following lysis, samples were either stored temporarily on ice prior to use in PCR, or kept at -20^{o}C for longer-term storage.

2.4.2 DNA probe hybridisation for species identification

Lyophilised primers (Table 6: Primer sequences used in this study.) were prepared by resuspending in 100µl of nuclease free water (master stock), and diluted 10x into separate tubes (Eppendorf) to create 10µM working stocks. Unless stated otherwise, PCRs were performed in 50µl volumes, according to Table 7.

Species	Gene	Primer sequence	Size of product
S. mitis	*pheA*-F	5`-TGGCTTATCCTTCCTAGATGG-3`	557bp
S. mitis	*pheA*-R	5`-GATTGCGGTCGACAAA-3`	
S. oralis	*rgg*-F	5`-GCTTTGACCGAACAGTTTCC-3`	475bp
S. oralis	*rgg*-R	5`-CATTGGTATTCCCCACCTTG-3`	
Anginosus group	*pbp2b*, MIL2B-F	5`-TGCTGCAACGGTAGCTAATGG-3`	275bp
Anginosus group	*pbp2b*, MIL2B-R	5`-CAAAGGTTTCTGCTGTCCCTG-3`	
S. pneumoniae	*cap3A*-F	5`-GAAACAGAGGTTAGGAAAGTAATC-3`	~1185bp
S. pneumoniae	*cap3A*-R	5`-ATATAACCGCCCAACGAATAA-3`	
S. pneumoniae	*cap3A*int-F (internal primer)-F	5`-TACTCCGACTAATTATGATGTAG-3`	~417bp
S. pneumoniae	*cap3A*int-R (internal primer)-R	5`-TTACCTCGCTATATGTATCTATCT-3`	

Table 6: Primer sequences used in this study. For *cap3A* and *cap3A*-int approximate sizes are given, as these regions are known to vary according to the number of tandem repeats contained within the amplified region (Waite et al., 2001).

PCRs were performed using a Verti, 96 well thermal cycler (Applied Biosystems®, Life Technologies)(WTSI and UoW) or a DNA engine Tetrad® II (Bio-Rad, Hercules, USA)(MLW). PCRs were undertaken with a denaturing step of 95°C for 30 seconds. Annealing temperatures were optimised between 50°C and 60°C, before an extension step was carried out at 72°C. The length of time for extension was calculated at 1 minute per kilobase of product length.

All PCR products were run on a 1% agarose gel, typically composed of 1g of DNA grade agarose (Invitrogen) and 100mL of TAE buffer (40mM Tris, 20mM glacial acetic acid, 1mM DTA, pH 8.0). This mix was superheated until the agarose dissolved, cooled to ~60°C and 5µl of ethidium bromide (10mg/mL) added. The agarose was poured into an appropriately sized gel tank, and allowed to set for ~1hr. TAE buffer was then added to submerge the gel and wells. Samples were mixed with loading dye according to the manufacturers instructions, and 10µl volumes loaded onto the gel. An appropriately sized DNA ladder was also added to determine product length. Gels were typically run at 120 volts and 250mA for 45 minutes. The PCR products were visualised by trans-illumination under ultraviolet light.

Volume	Reagent
34.8µl	sterile distilled water
5µl	10X PCR buffer (500 mM KCl, 100 mM Tris-HCl (pH 9.0), 1.0% Triton X 100)
3µl	25 mM $MgCl_2$
1µl	10 mM dNTPs (10 mM each dATP, dTTP, dGTP. dCTP)
2µl	10 µM forward primer
2µl	10 µM reverse primer
0.2µl	Taq polymerase
2µl	Template
50µl total volume	

Table 7: PCR constituents and volumes used in this study.

2.5 DNA extraction

2.5.1 Phenol Choloroform extractions

Cells were grown on blood agar plates overnight in combination with an optochin disc (Oxoid), and checked for contamination prior to use. Broth cultures were prepared by inoculating 10ml of BHI broth (Oxoid) with a single colony picked using a sterile loop. Broths were cultured at 37°C at 5% CO_2 for 18-20 hours, before being centrifuged at 4000rpm for 10 minutes, and the supernatant removed to leave a bacterial pellet.

Alternatively, a plate culture was prepared by inoculating one or more blood agar plate with a single pneumococcal colony, and incubated overnight. A 1mL drop of PBS was deposited onto the surface of the plate following growth, and a spreader used to remove the bacterial growth, which was removed using a pipette into a 2µL tube (Eppendorf), and centrifuged at 8000rpm for 10minutes. The supernatant was then removed to leave a bacterial pellet. Plate culture was used where possible in order to check the colony morphology of pneumococci prior to sequencing.

Bacterial pellets were then transferred to a 2mL tube (Eppendorf) and resuspended in 1mL of 25% sucrose-Tris buffer (pH8.0)(Invitrogen). 100µl of lysozyme was added to the resuspended pellet, and the sample vortexed for 10

seconds. The sample was then transferred to a heat-block set at 37°C for 15 minutes. Following incubation, 50µl of proteinase K (Qiagen, Venlo, Netherlands), 30µl of RNase A (Qiagen) and 650µl NP40 lysis buffer (Thermo Fisher Scientific)(or 250 sarkosyl + 400µl EDTA) was added. The resulting solution was incubated in a heat-block set at 50°C for between two and 4 hours, or until the suspension cleared. Eppendorf tubes, containing the lysed cells were then allowed to cool to room temperature before the lysate was transferred to a 15mL Falcon™ tube (Thermo Fisher Scientific) and the contents made up to a final volume of 2.5mL with nuclease free water. 2.5mL of phenol-chloroform (phenol:chloroform:isoamyl alcohol, 25:24:1, pH8)(Thermo Fisher Scientific) was then added to the tube. The closed tube was gently inverted several times to mix the sample, but limit DNA shearing, so that a white emulsion became apparent. The tube was then centrifuged at 4000rpm at 4°C for 10 minutes, so that the emulsion separated. Samples were spun for more time (e.g. 30 minutes) if the emulsion failed to separate initially. In the fume hood, the top aqueous phase was transferred to a fresh Falcon™ tube, containing 2.5mL of phenol-chloroform. During this stage it is important not to transfer any of the interface between the two phases, as this leads to contamination of the sample being removed. Equally it is important that as much of the aqueous phase is recovered as possible. The Falcon™ tube containing the recovered aqueous phase was once again inverted, and centrifuged in order to separate the phases. This process was repeated up to 3 times, or until the debris had visibly cleared from the interface between the two phases. At the last round of cleaning, 2.5mL of chloroform (Sigma Aldrich Inc) alone was added to the recovered phase, prior to centrifugation and recovery. On recovery of the aqueous phase for the final time, the phase was transferred to a Falcon™ tube and 5mL of 100% ethanol added. This mixture was incubated for a minimum of 4 hours, or overnight at -20°C.

Following incubation, the sample was then centrifuged for 30 minutes at 4000rpm and 4°C. The Falcon™ tube was then smoothly inverted to remove the ethanol, while taking care not to disturb the pelleted DNA. 5mL of ice cold 70% ethanol was then added, and the sample centrifuged for a further 30 minutes at 4000rpm and 4°C. The ethanol was then discarded as above, and the pelleted

sample left to air-dry. Following drying, the pellet was resuspended in 100-500µl of TE buffer or nuclease free water (DNA rehydration solution) and stored at 4°C.

2.5.2 Ethanol precipitation

Ethanol precipitation can be used to increase the purity and concentrate DNA following extraction, such as was described in the above phenol-chloroform method. The function of this method was to precipitate the negatively charged nucleic acids from a solution, through the addition of ethanol and a salt (such as sodium cations) (Zumbo, 2012).

During the recovery of PCR amlicons for sequencing, the PCR mix was added to 2/3 volume of sodium acetate, and twice this volume of 100% ethanol. This mixture was inverted several times, and incubated overnight at -20°C. Eppendorf volumes (<2ml) can be centrifuged at max speed at 4°C for 30 minutes; otherwise, the sample was centrifuged at 4000rpm for 30 minutes at 4°C. Excess ethanol was removed by inversion, taking care to leave the DNA pellet undisturbed. 1mL of ice cold 70% ethanol was then added to the pellet, which was centrifuged for a further 30 minutes. The ethanol was then removed by pipette and the DNA pellet allowed to air-dry before being resuspended in TE buffer or nuclease free water (DNA rehydration solution) and stored at 4°C until use.

2.6 Quantification

DNA samples were quantified using a NanoDrop® ND-1000 (Thermo Fisher Scientfic) or Qubit® Fluorometer (Invitrogen). For sequencing on an Illumina MiSeq (Illumina Inc, San Diego, USA) it was specified that samples be composed of >1µg RNA free, double-stranded DNA, dissolved in 10mM Tris pH8, 0.1mM EDTA at a concentration of between 50 and 150ng/µL.

2.6.1 NanoDrop® ND-1000

The NanoDrop® system (Thermo Fisher Scientific) allows for the fluorescent measurement of DNA concentrations independent of the need for cuvettes or capillaries, as the sample is loaded directly into the apparatus (Figure 23) (Desjardins and Conklin, 2010). The upper and lower optical surfaces, onto which the sample was loaded was cleaned prior to use by loading 1-2µl of deionised water onto the lower optical surface and closing the lever arm so that upper and lower surfaces touched. Opening the lever arm, both upper and lower optical surfaces were wiped prior to use. Opening the software for the NanoDrop® ND-1000, the module for measuring nucleic acids was selected. The spectrophotometer was initialised by placing 1µl of clean water onto the lower optical surface, closing the lever arm, and selecting the "initialise" option on the NanoDrop® software. Initialisation took ~10 seconds, and once completed a Kimwipe® was used to clean the upper and lower optical surfaces once again. A blank measurement was then taken by loading 1µl of DNA rehydration solution onto the optical surface, the lever arm closed, and "blank" selected from the NanoDrop® software. This provided a baseline reading, which was subtracted from all subsequent measurements. If the measurement of DNA rehydration solution was not zero or was not very close to zero, the blank was repeated. Each sample was then read by loading 1µl of sample onto the lower optical surface, the lever arm being closed, and the "measure" option chosen from the software. A Kimwipe® was used to clean the upper and lower optical surfaces between readings, and a blank was retaken every five DNA sample measurements.

At each reading the following output was given:

- A predicted DNA concentration in ng/µl
- A 260/280 absorbance ratio, which was expected to be in the range ~1.8nm. Lower values than this indicated the presence of protein, phenol or contaminants absorbing closer to 280nm
- A 260/230 ratio, which was expected to be in the range 2.0-2.2nm. Ratios lower than this indicated contaminants absorbing closer to the 230nm spectrum.

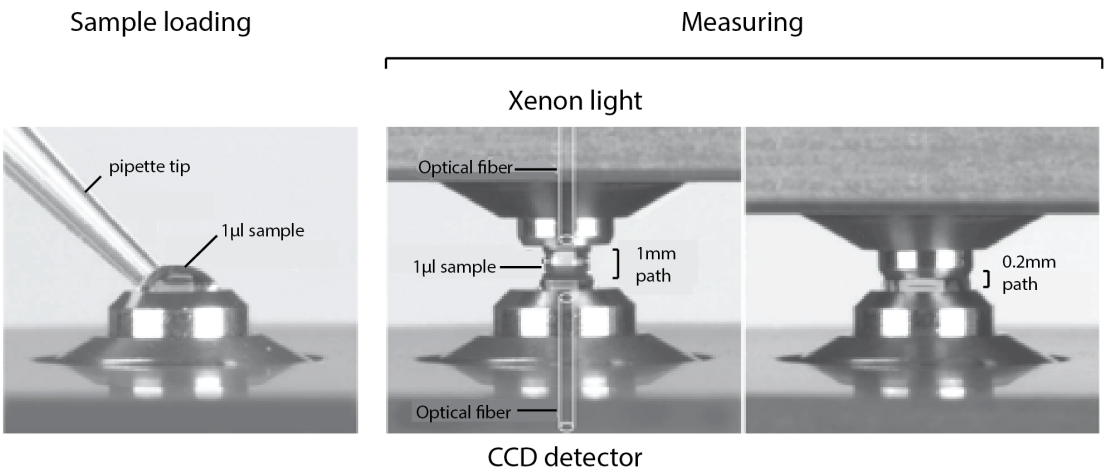


Figure 23: The NanoDrop® requires 1µL of sample to be loaded onto the lower pedestal, prior to closing of the lever arm. This creates a liquid column, through which two readings are taken at different distances to aid nucleic acid detection (source: http://openwetware.org/wiki/BISC209/F13:_Lab5).

2.6.2 Qubit® Fluorometer

The Qubit® Fluorometer (Invitrogen) is an alternative apparatus used for the measurement of DNA (Figure 24). The Qubit® Fluorometer uses fluorescence based dyes, which bind specifically to DNA, RNA, or protein contrasting the NanoDrop® method which is based on UV absorbance. Consequently the Qubit® Fluorometer can measure both RNA and DNA and is more sensitive at detecting lower concentrations of material. The Qubit® Fluorometer however is unable to detect contamination (Life Technologies, 2015). All reagents used for Qubit® Fluorometer (Invitrogen) readings were included in the Invitrogen dsDNA HS Assay kit, unless stated otherwise.

A working solution was first prepared, by diluting Qubit® reagent with Qubit® buffer at a ratio of 1:200. The volume of working solution required was calculated by: 200µL x [the number of samples to be tested + 2].

Two standards for instrument calibration were prepared by adding 10µL of either Standard No. 1 solution (10ng/µL) or Standard No. 2 solution (10ng/µL) to a 0.5mL PCR tube (Eppendorf) containing 190µL of working solution.

In addition, for each of the samples measured, 1µL of sample was added to 199µL of working solution in a 0.5mL PCR tube (Eppendorf). Once all of the

samples and standards had been prepared, these were vortexed for 2-3 seconds, and left to incubate for 2 minutes at room temperature, ensuring any bubbles were removed prior to use.

The fluorometer was then calibrated by inserting Standard 1 into the sample chamber, taking a calibration reading, and repeating this for Standard 2. Each sample was now measured, and re-calibration with standards repeated every five sample readings. The concentration of DNA in the original sample was then calculated from the fluorometer output as follows:

Sample DNA concentration = Qubit® value x measured volume (200µL)/sample volume added (1µL)

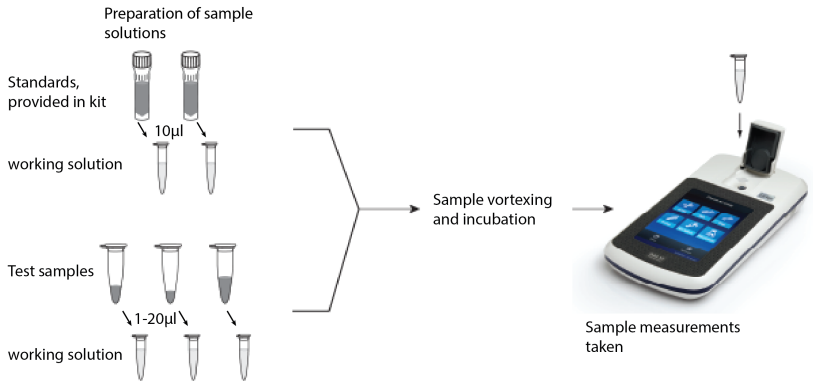


Figure 24: Sample readings taken using the Qubit® fluorometer (source http://tools.thermofisher.com/content/sfs/prodImages/high/Qubit-3.0-3qrtr-RightClosed-7in72dpi_33216.jpg).

2.7 Sequencing

Sequencing was performed by the core sequencing team at the WTSI. All large datasets were sequenced using the Genome Analyzer II (Illumina) in multiplexed libraries compiled from up to 12 end-tagged-isolate libraries at a time. NGS libraries were prepared from cleaved genomic DNA, which was ligated to adapters, and DNA constructs of ~250bp in length were selected. These were subjected to 18 cycles of PCR, and the reverse strands chemically removed, prior to being loaded onto the single lane of an Illumina Genome Analyser II flow cell (Illumina) for sequencing.

Mutant isolates and controls used in chapter 6 were sequenced using the Illumina MiSeq, and samples prepared similar to that described above. Capillary sequencing (Applied Biosystems 3730xl platform) was used for sequencing of the *cap3A* gene.

2.7.1 *De novo* Genome Assembly

De novo assemblies were constructed using an in-house wrapper script developed at the WTSI, using the program Velvet (www.ebi.ac.uk/~zerbino/velvet/) (Zerbino and Birney, 2008). The reads for each isolates are first split into k-mers, shorter sequences, varying in length between 66% and 90% of the total read length. K-mers allow reads of equal length to be generated (raw reads varying in length by around 100bp), and are necessary in order to fulfil the assumptions de Bruijn graph theory- principally that a sequence of length L must overlap an adjoining sequence by length L-1. As such, all possible k-mers must be present for graph construction. Low coverage spots, and sequencing errors mean that this is not always possible using the raw data, and hence short reads are split into shorter k-mers until this can be achieved. For each k-mer size, a de Bruijn graph is used to generate an assembly, represented by a, or a series of contigs (Figure 25). These assemblies are scored, based on identifying the size at which half of all assembled bases reside in a contig of a given size or longer, termed a N50 score (Martin and Wang, 2011). The contig assembly with the highest N50 score is then selected. In addition contigs whose length is shorter than the length of the insert size are thought to represent spurious contigs and are removed. Following this, an assembly improvement step is run. Longer scaffolds are generated during this process by iteratively running SSPACE (SSAKE-based Scaffolding of Pre-Assembled Contigs after Extension) (Boetzer and Pirovano, 2014), which uses order, distance and orientation information about contigs generated from paired-end read assemblies. 120 iterations of the program GapFiller (Boetzer and Pirovano, 2012) are used to close scaffold gaps between contigs. GapFiller identifies read pairs in which one member of the pair maps to the contig, and the second fails (partially) within the gap. An attempt is then made to use these unmapped reads

to close any gaps, based on identifying overlapping k-mers. Finally all reads are re-aligned against this assembly, using SMALT (WTSI). SMALT generates an index of sequences, <21 nucleotides long, sampled at equal distances along the length of the reference sequence (termed a hash index). By identifying matches between short reads and the sequences in the hash index (seed matches), reads are subsequently aligned using a local sequence alignment algorithm (Smith-Waterman algorithm).

Where complete annotated assemblies were not available, contigs from the above mapping pipeline were ordered using ABACAS (abacas.sourceforge.net), and the Rapid Annotation Transfer Tool (RATT)(ratt.sourceforge.net) (Otto et al., 2011) used to transfer annotations from a closely related assembled sequence. The program ACT was then used to compare *de novo* assemblies to previously annotated references to check for annotation and assembly errors. The newly constructed and annotated reference could then be used to map sequenced isolates.

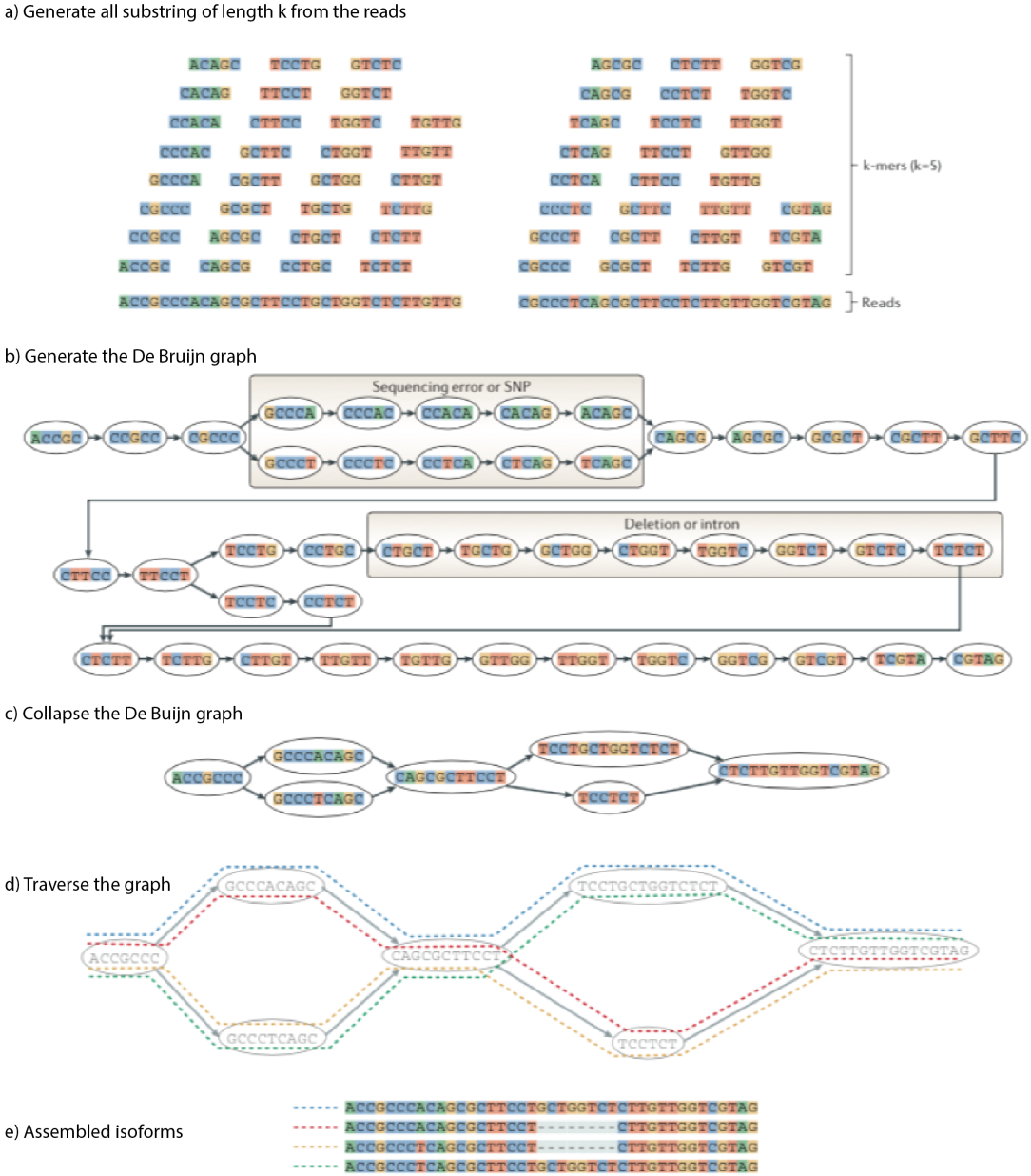


Figure 25: *De novo* read assembly. Each read is first split into substrings (k-mers)(a), which are used to represent nodes on a de Bruijn graph, with overlapping k-mers (length of read -1) being joined. Sequencing errors, and indels are then represented by bulges in the graph (b). Continuous overlapping k-mers are joined to collapse the graph, traversing the graph by all possible paths (d) then allows isoforms to be assembled (Martin and Wang, 2011)

2.7.2 Mapping pipeline

Reference based read assembly was performed using the WTSI in-house pipeline. By default reads were aligned using the hash based SMALT aligner, described above. The Burrows-Wheeler Aligner (BWA) was also used where stated, which uses a BW based algorithm, rather than hashing for library construction (Li and Durbin, 2009, Langmead et al., 2009).

Both read alignment programs created BAM (Binary Alignment Map) files, which could be viewed in conjunction with the reference sequence in Artemis to check for assembly errors.

2.7.3 Bioinformatic analyses

Indels (insertions/deletions) were called using a pipeline developed by Dr S. Harris at the WTSI (Harris et al., 2013b). This pipeline first identifies candidate indels using the program PINDEL (Ye et al., 2009). Predicted indels are then subject to quality controlling, by scoring mappings produced in the presence and absence of that indel. Where the inclusion of an indel is found to improve the quality of the mapping, it is included. The program LASER, part of the MateClever toolkit (Marschall et al., 2013) was used as an alternative method of indel detection, run in conjunction with the program BWA (Li and Durbin, 2009). LASER (Marschall et al., 2013) and PINDEL (Ye et al., 2009) are discussed in more detail in chapter 7.

SNPs were identified using the method described in the supplementary information of Harris and colleagues (2013a) from the raw assembly. Briefly, the mapping quality score, a measure of the probability that a read is misplaced, and the phred score, which is assigned to each base called in the trace file, were considered. Bases with a mapping quality score above 30 and a phred score above 50, (giving a theoretical accuracy of 99.999% (www.phrap.org; www.illumina.com)), that differed from the reference sequence were compared to the other reads mapped at that location. Where the SNP was supported by 75% of at least 4 mapped reads, with at least two of these reads mapping to the forward and reverse strands, the SNP was included. Otherwise the position was marked as unknown, “N”.

2.7.4 *In silico* serotyping and sequence typing

In silico serotype detection was performed using a WTSI in-house script written by Dr Nicholas Croucher. The script counts the number of input reads (Illumina or 454) that are aligned to a reference using a BWA aligner. By mapping input reads against each of the 91 pneumococcal capsule synthesis genes (Bentley et

al., 2006), the script assigns a serotype based on the highest scoring mapping. Sequence typing was performed using a similar method, substituting the reference sequences with pneumococcal sequence types.

In addition, the program RAxML (Stamatakis et al., 2005) was used for Maximum-Likelihood phylogenetic tree construction, and bespoke sequence alignments were performed using MUSCLE (Edgar, 2004). Recombination events were identified based on SNP divergent clusters using a WTSI in-house script (Croucher et al., 2015). Sequence alignments were viewed using SEAVIEW Multiple sequence alignments were viewed in SEAVIEW (http://doua.prabi.fr/software/seaview), and promoter regions were identified using PEPPER (www. pepper.molgenrug.nl). Genome comparison was performed using BLAST (Altschul et al., 1997), and visualisation between sequences was performed using ACT (Carver et al., 2005).

Additional scripts developed in the subsequent analyses were written in Python™ version 2.7 (Python Software Foundation, 2013), or R, version 3.0.2 (R Development Core Team, 2013).

3 Mechanisms of Beta-lactam Resistance

3.1 Introduction

The beta-lactams are a widely available class of antimicrobials and have a strong efficacy in the treatment of pneumococcal disease. As such, for resource poor countries such as Malawi, beta-lactams are often favoured for treating pneumococcal disease. The spread of beta-lactam resistance worldwide (see introduction) whilst concerning for developed countries may present a greater challenge for countries that are already restricted in their access to medical supplies and treatments. Malawi is currently reliant on only two beta-lactams, ceftriaxone and penicillin for treating pneumococcal disease. However resistance to both of these antimicrobials is emerging in Malawi (Everett et al., 2011), and has been recorded globally (Appelbaum, 1987).

A long history of study into the development of beta-lactam resistance among pneumococci reveals that there are many paths by which resistance can develop. However, most routes to resistance are associated with fitness costs, especially as beta-lactams target a particularly important and sensitive pathway within the bacterium. Epidemiologically therefore, dominant bacterial lineages that have gained resistance without any measurable fitness cost are particularly of interest for controlling pneumococcal disease.

Pneumococcal populations are frequently dominated by a small number of lineages. The Pneumococcal Molecular Epidemiology Network (PMEN) was established in order to better study such dominant lineages, many of which are also associated with MDR. The first incidence of PMEN lineages within Malawi was reported in 2012, where three PMEN lineages (PMEN 19 ST289 serotype 5, 25 ST63 serotype 14, 27 ST217 serotype 1) were identified. All of these currently circulating PMEN lineages were found to be covered by the recently introduced PCV13 (Everett et al., 2012). However, whilst PCV13 is likely to help reduce the burden of pneumococcal disease in the short term, serotype switching, and the introduction of other PMEN lineages is likely to occur in the future.

3.1.1 Current understanding of beta-lactam targets

PBP enzymes play an essential role in the maintenance of the pneumococcal cell wall by crosslinking peptidoglycan polymers (Park and Uehara, 2008). Beta-lactams, which possess a spherical amide ring, share a structural similarity to D-alanyl-D-alanine dipeptides, the natural substrate for PBP transglycosylation activity (Figure 26)(Tipper and Strominger, 1965). Transglycosylation of D-alanyl-D-alanine side branches carried on adjacent glycan strands forms the cross-linking within the cell wall. When transglycosylation uses beta-lactam as a substrate, hydrolysis and release of the acylated PBP is prevented, as the beta-lactam assumes an open ring structure on binding, so that subsequent hydrolysis of the acyl bond, and release of the beta-lactam, cannot occur (Kohanski et al., 2010).

Although the pneumococcus carries a complement of 6 PBP enzymes, only modification of PBPs 1a, 2b and 2x are thought important for beta-lactam resistance. All three enzymes share a common transpeptidation activity, with PBP2b thought to play a wider role in cell elongation (Berg et al., 2013), and PBP2x in cell division (Perez-Nunez et al., 2011, Land et al., 2013, Tsui et al., 2014). Modifications that affect the transpeptidation domains of these enzymes are thought particularly important for the development of resistance, as these will affect the ability for the antibiotic to compete with the native D-alanyl-D-alanine substrate.

Whereas beta-lactams are effective against gram-positive and gram-negative bacteria, glycopeptides, which similarly inhibit of cell wall synthesis, are only effective against gram positive bacteria, which have a greater cell wall permeability compared to gram negatives (Lambert, 2002).

Cell wall inhibiting enzymes are thought to exert their lethality through both lysis dependent, and lysis independent routes. The lysis dependent route remains to be fully understood (Tomasz, 1979), but appears to be linked to amidase activity (Tomasz et al., 1970, Kohanski et al., 2010). Isolates deficient in amidase activity (via peptidoglycan hydrolases, or autolysins) are found to be

tolerant (survival but not growth), of otherwise lethal antimicrobial dosages (Kohanski et al., 2010). As such beta-lactams appear to exert their lethality through both the inhibition and active degradation of peptidoglycan (Tomasz et al., 1970). Lysis independent death occurs at a slower rate, with mechanisms of regulation by two component systems having so far been described. Regulation of the *S. pneumoniae* autolysin LytA of by the VncSR two-component system is found to be involved in lysis-dependent (Novak et al., 1999) and independent (Novak et al., 2000) tolerance to penicillin and vancomycin (Kohanski et al., 2010). A link between SOS activation, via recA, and beta-lactam inhibition has also been suggested to occur in some bacteria. Here beta-lactams disrupt central metabolism, particularly NADH consumption resulting hydroxyl formation, and DNA damage, triggering the SOS response (Kohanski et al., 2007, Kohanski et al., 2010).

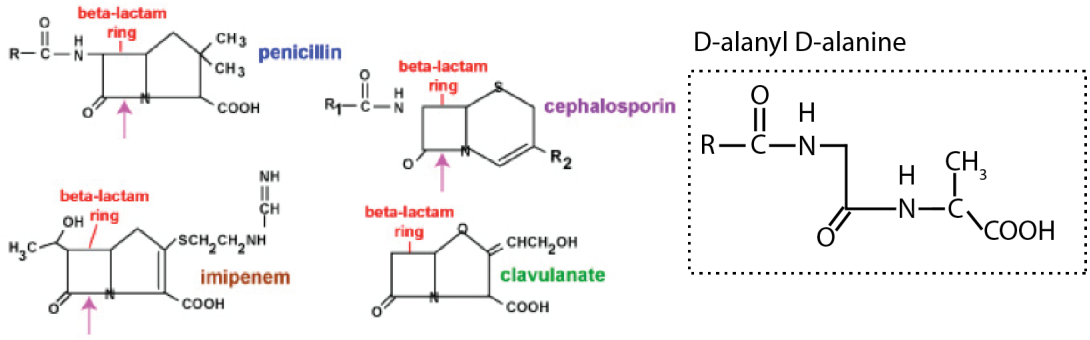


Figure 26: Chemical structures of different beta-lactam classes. The D-alanyl D-alanine structure to which beta-lactams acts as substrate mimics is included for comparison (source: http://faculty.ccbcmd.edu/courses/bio141/lecguide/unit2/control/images/betalactam.jpg).

3.1.2 Beta-lactam antibiotics: Overview

Beta-lactam antibiotics share a common structure of a beta-lactam ring, attached to a variable side-chain (Miller et al., 2001). The group is comprised of a number of antibiotic classes, which include penicillins, the cephalosporins, the penems, carbapenems, monobactams, and combination therapies (Figure 26). Glycopeptides act by inhibiting elongation of peptidoglycan, which vancomycin achieves by binding D-alanyl-D-alanine, inhibiting transglyosylase activity (Kohanski et al., 2010). Synthetic derivatives of vancomycin, which bind sites other than D-alanyl-D-alanine, but similarly inhibiting cell wall synthesis have also been identified (Ge et al., 1999). Restricted usage of vancomycin has meant

that resistance to this antimicrobial is far more limited than compared to the beta-lactams (Novak et al., 1999, Moellering, 2006).

3.1.3 Penicillins

Penicillins have a thiazoldine ring attached to a beta-lactam ring. Natural penicillins are widely effective against gram-positive aerobic bacteria, such as the streptococci, enterococci and non-beta-lactamase staphylococci (Miller et al., 2001). The newer synthetic penicillins have an extended spectrum of activity to include *H. influenzae, N. gonorrhoeae* and *E. coli* (Miller et al., 2001). The coupling of a beta-lactam to a beta-lactamase inhibitor can further increase the spectrum of activity, although pneumococci lack such plasmid encoded resistance mechanisms. The penicillins include benzyl-penicillin, colloquially penicillin, which was the first of this class of antimicrobials to be identified. They are bactericidal to actively growing and dividing bacteria.

Penicillin susceptible pneumococcal isolates carry an MIC of 0.0008μg/mL. Deviations above this value therefore suggest some degree of PBP modification (Dowson et al., 1989). Isotopic labelling was initially used to demonstrate how different classes of beta-lactams show particular affinities for sets of PBP enzymes (Munoz et al., 1992). Significantly, this indicated how cephalosporin resistance could be achieved through the replacement of susceptible forms of *pbp2x* and *pbp1a* with alleles from resistant isolates, achieving intermediate-resistance and full resistance respectively (Munoz et al., 1992). In contrast, penicillin resistance has been shown to arise through the sequential modification of *pbp2x*, *pbp2b* and *pbp1a*- the latter necessary for high-level resistance (Smith and Klugman, 1998).

The availability of nucleotide sequencing has since led to a focus on identifying the SNPs responsible for resistance. The interaction between substrate and three conserved motifs within the transpeptidase domain has been identified as key to the PBP-beta-lactam interaction (Sauvage et al., 2008). However, despite a number of studies attempting to identify the pathways by which beta-lactam

resistance occurs using laboratory induced mutation (Guenzi et al., 1994, Grebe and Hakenbeck, 1996), this has not been found to reflect the *pbp* variation observed in clinically resistant isolates. As such there has been an increasing shift from laboratory induced resistance, to the study of resistance in clinical isolates (Smith and Klugman, 1998, du Plessis et al., 1998).

This approach has led to a number of attempts at characterising the SNPs present in clinically resistant isolates (Granger et al., 2005), however few studies have successfully proven the ability of a particular SNP to transform a susceptible isolate to resistance (Smith and Klugman, 1998, du Plessis et al., 1998).

The ability to prove the link between SNP and phenotype is complicated by two processes. Firstly beta-lactam resistance arises in pneumococci through recombination between divergent bacteria, rather than through *de novo* mutation. This results in the occurrence of gene mosaics, with sections of highly divergent sequence being present within the gene, that have originated from different bacterial species (Dowson et al., 1989). As such it can be unclear how much of this variation actually modifies an isolate's beta-lactam susceptibility, and how much is simply an artefact of frequent recombination within these genes.

A second complication stems from the observation that identical *pbp* alleles can be present in isolates of different beta-lactam susceptibilities (Chesnel et al., 2005). This indicates that interactions are also occurring between *pbp* alleles and the genetic background, in addition to the interactions between *pbp* genes (Orio et al., 2011). The traditional approach of studying *pbp* alleles in isolation will consequently fail to identify these additional mechanisms of resistance.

Modes of beta-lactam resistance independent of *pbp* alleles have been identified. Differences in the ability of MurM to alanylate rather than serylate the growing stem peptide at the lysine residue appears to alter the competitive balance between beta-lactams and stem peptides for the transpeptidation domain (Smith

and Klugman, 2001, Lloyd et al., 2008). The involvement of genes independent of those conventionally involved in peptidoglycan synthesis has also been suggested, such as the putative glyocosyltransferase (CpoA) and a histidine protein kinase (CiaH) (Hakenbeck et al., 1999). The ability for processes prior to the interaction between PBPs and beta-lactams to alter the phenotypic level of resistance means that there is a potential for a much larger number of genes to be involved in resistance. Importantly, such modification could occur in genes less critical for survival than the *pbps*, facilitating the maintenance and spread of such resistance. The ability to identify such candidates *in vitro* is however limited by throughput, and a lack of methods to identify gene candidates for hypothesis testing.

The increasing availability of whole genome sequences for bacteria has led to a new field of comparative genomics. As such it is now possible to view the genetic variation present within and across multiple genomes. This has allowed the development of tools to characterise the genetic processes underlying genetic variation across bacterial populations. There is now an increasing drive to link bacterial phenotypic variation to nucleotide variation identified *in silico*.

3.1.4 Aims

- To characterise the occurrence and conservation of SNPs known to be associated with beta-lactam resistance in PMEN1
- To identify SNPs associated with changes in beta-lactam susceptibility and to use this information to identify gene candidates, previously unknown to affect beta-lactam susceptibility

3.2 Study Isolates

PMEN1 is a highly successful MDR pneumococcal lineage, characterised as carrying a 23F serotype, and ST81. It is an example of a lineage that is paradoxically successful despite the heavy mosaicism within its *pbp* alleles (Coffey et al., 1991). The first PMEN1 strain was isolated from Barcelona, Spain, in 1984, although recent phylogeny reconstructions suggest that the lineage

originated in approximately 1970 (Croucher et al., 2011). The lineage rapidly spread from this single source, to become dominant worldwide throughout the 1980's. Its spread was characterised by a high association with MDR and IPD (Munoz et al., 1991). For this reason, the PMEN1 group are widely regarded as the most "successful" of the described PMEN lineages- success being measured in terms of its global prevalence. The isolates used in this study represent a set of 173 globally sampled PMEN1 isolates, from a mixture of carriage and IPD, collected globally. The dataset was collated previously (Croucher et al., 2011), and has been well characterised *in silico* in terms of nucleotide polymorphisms, and recombination events within isolates and across the dataset. In addition a single strain, isolated prior to the identification of PMEN1 was used for context and comparison.

3.3 Characterising Beta-lactam Resistance in PMEN1

3.3.1 Reference sequence and conserved resistance SNPs

The long history of study into penicillin resistance has meant that pbp genes have been well described in resistant isolates. A number of polymorphisms have been associated, and in some cases shown to alter beta-lactam susceptibility. To determine the occurrence of such mutations in the PMEN1 isolates, a read alignment file was generated for *pbp1a*, *pbp2b*, and *pbp2x* alleles. Nucleotide polymorphisms were then identified relative to *pbp* genes present in the reference strain R6, which was isolated prior to the introduction of penicillin, and as such is regarded as universally beta-lactam susceptible (Griffith, 1928, Lederberg and Gotschlich, 2005). This analysis allowed the characterisation of mutations present across the PMEN1 dataset. In addition, it was possible to determine whether these were identified, based on the literature, as associating with changes in beta-lactam susceptibility. Furthermore, it was possible to determine whether these occurred in earliest PMEN1 isolates (ATCC700669), or whether these occurred subsequent to the emergence of this MDR clone.

Given the focus on penicillin resistance in the literature, most of the SNPs discussed herein correspond to resistance to this class of antibiotic.

3.4 Characterising PMEN1 resistance mutations

3.4.1 *Pbp1a*

Pbp1a modification is thought necessary for the development of high-level resistance to penicillins and cephalosporins. In accordance with this several mutations that have previously been proven to associate with a loss in susceptibility to beta-lactams were identified (Table 8). Furthermore, these SNPs were conserved across the PMEN1 lineage in all but one case. Of note was the T371A mutation present in the STMK amino acid motif, previously found to result in a change in penicillin susceptibility from 0.5μg/mL to 1.5μg/mL (Smith and Klugman, 1998), consistent with *pbp1a* modifications being necessary for higher-level penicillin resistance. SNPs identified by Granger and colleagues (2005) were not tested for phenotypic effects, but were identified within other resistant lineages present in Quebec, Canada. Consequently, whilst a number of SNPs present corresponded to those identified in their study (Granger et al., 2005), such associations could be an artefact of recombination, and may not affect the MIC of the strain.

pbp1a	Confirmed association?	Conserved	Drug tested	Reference
T371A	Yes	100%	Penicillin	(Smith and Klugman, 1998)
I393M	No	100%	Penicillin	(Granger et al., 2005)
H395N	No	100%	Penicillin	(Granger et al., 2005)
E397I	No	100%	Penicillin	(Granger et al., 2005)
N405S	No	100%	Penicillin	(Granger et al., 2005)
P432T	No	100%	Penicillin	(Granger et al., 2005)
I459M	Yes	100%	Penicillin	(du Plessis et al., 1999)
S462A	Yes	100%	Penicillin	(du Plessis et al., 1999)
T574N	Yes	100%	Penicillin	(Smith and Klugman, 1998)
S575T	Yes	100%	Penicillin	(Smith and Klugman, 1998)
Q576G	Yes	99% (absent in 3 isolates)	Penicillin	(Smith and Klugman, 1998)
F577Y	Yes	100%	Penicillin	(Smith and Klugman, 1998)
L583M	No	100%	Penicillin	(Smith and Klugman, 1998)
A585V	No	100%	Penicillin	(Smith and Klugman, 1998)
N609D	No	100%	Penicillin	(Dias et al., 2009)*

Table 8: SNPs found within ATCC700669, the reference strain for PMEN1. *Uses a member of the Spain^{23F}-1 clone, identifying this SNP as commonly associated with beta-lactam resistance.

3.4.2 *Pbp2b*

Although often highly modified in penicillin resistant isolates, and required for the early development of beta-lactam resistance, only a single SNP in the PMEN1 *pbp2b* allele was identified in the literature as being associated with a loss in beta-lactam susceptibility (Table 9). The T446A mutation close to the SSN motif was however conserved across all the isolates in this collection, and has been found to result in a 60% reduction in penicillin sensitivity (Pagliero et al., 2004).

pbp2b	Confirmed association?	Conserved	Drug tested	Reference
T446A	Yes	100%	Penicillin	(Pagliero et al., 2004)

Table 9: SNPs found in reference PMEN1 reference ATCC700669, and prevalence among other PMEN1 isolates of the dataset.

3.4.3 *Pbp2x*

Pbp2x similar to *pbp2b* is often highly modified in resistant pneumococci, corresponding with the development of low-level beta-lactam resistance. *Pbp2x* possessed the greatest number of resistance associated SNPs, although few of these had been proven to reduce beta-lactam susceptibility (Table 10). Of note was the well-characterised T338A mutation, which occurs within the SxxK motif (Zerfass et al., 2009). In addition an I371T mutation, which also occurs in highly resistant *S. mitis* was also conserved across PMEN1 isolates. Many of the SNPs identified in *pbp2x* were found to occur in clinical isolates resistant to both penicillin and cephalosporins.

pbp2x	Confirmed association?	Conserved	Drug tested	Reference
Q281L	Yes	100%(1 isolate)	Penicillin	(Q281P) (Fani et al., 2011)
T338A	Yes	100%(1 isolate)	Penicillin	(Zerfass et al., 2009)
I371T	Yes	100%(1 isolate)	Penicillin, cephalosporins	(Amoroso et al., 2001) (from *S. mitis*)
R384G	Yes	100%(1 isolate)	Penicillin, cephalosporins	(Carapito et al., 2006)
S389L	Yes	100%(1 isolate)	Penicillin	(Dessen et al., 2001)
N417K	No	100%	Penicillin	(Granger et al., 2005)
I462L	No	100%	Penicillin, cephalosporin	(Pernot et al., 2004)
A491V	No	100%	Penicillin, cephalosporin	(Pernot et al., 2004)
D506E	No	100%	Penicillin, cephalosporin	(Pernot et al., 2004)
N514H	Yes	100%(1 isolate)	Penicillin, cephalosporin	(Dessen et al., 2001)
V516I	No	100%	Penicillin, cephalosporin	(Pernot et al., 2004)
V523L	No	100%	Penicillin, cephalosporin	(Pernot et al., 2004)
S531Y	No	100%	Penicillin, cephalosporin	(Pernot et al., 2004)
T536I	No	100%	Penicillin, cephalosporin	(Pernot et al., 2004)
L546V	Yes	100%(1 isolate)	Penicillin	(Ho et al., 2004)
L565S	No	99% (2 isolates)	Penicillin, cephalosporin	(Chi et al., 2007)
D567N	Yes	100%	Penicillin, cephalosporin	(Carapito et al., 2006)
N605T	Yes	100%		(Sadowy et al., 2010b)
L710F	No	100%	Penicillin, cephalosporin	(Pernot et al., 2004)
Q721E	No	100%	Penicillin, cephalosporin	(Pernot et al., 2004)
T745K	No	100% (1 isolate)	Penicillin, cephalosporin	(Pernot et al., 2004)

Table 10: SNPs found in reference ATCC700669, and prevalence among other members of the dataset.

3.5 Antibiotic Susceptibility Testing

Antibiotic susceptibilities across the PMEN1 lineage were tested using antibiotic impregnated plates and a Steller's replicator, allowing 48 individual strains to be

tested in parallel (see methods). Antibiotic MICs were determined in triplicate for all strains. In cases where results were not consistent, the highest MIC was recorded. A set of seven beta-lactam antibiotics was chosen based on recommendations for treating pneumococcal diseases (CDC recommendations)(Table 11), in addition, the glycopeptide vancomycin was included as this antimicrobial also targets the cell wall synthesis pathway.

Unlike many pneumococcal lineages, the association between PMEN1 and the carriage of MDR meant that all strains carried some degree of altered antibiotic susceptibility. As such for each antibiotic, a suitable antibiotic range over which to test MICs was first identified. MIC ranges were chosen by identifying antibiotic concentrations that delimited the PMEN1 isolates into three or more groups. Consequently, for some antibiotics, non-standard MIC values were chosen (i.e. concentrations were not doubled). Antibiotic susceptibility testing was performed in accordance with the guidelines given by EUCAST for dilution testing.

Antibiotic	Recommended For	Class	Other groupings
Vancomycin	Pneumococcal meningitis	Glycopeptide	
Benzyl penicillin (colloquially "penicillin")	CAP	Penicillins (penams), narrow spectrum (beta-lactamase sensitive)	Natural penicillin
Amoxicillin	CAP, otitis media, sinusitis	Penicillins (penams), extended spectrum	Aminopenicillin (can be combined with beta-lactamase)
Ampicillin	CAP, meningitis	Penicillins (penams), extended spectrum	Aminopenicillin (can be combined with beta-lactamase)
Ceftriaxone	CAP, meningitis	Cephalosporins (cephems), 3rd generation	
Cefotaxime	CAP	Ceohalosporins (antipseudomonal), 3rd generation	
Ceftazidime	CAP	Cephalosporins (antipseudomonal), 3rd generation	
Cefalexin	CAP	Cephalosporins, 1st generation	

Table 11: Antibiotics recommended for treatment of pneumococcal disease, based on CDC recommendations. CAP: Community Acquired Pneumonia, IPD: Invasive Pneumococcal Disease.

3.5.1 Penicillin and Cephalosporin Susceptibilities

Figure 27 shows the antibiotic susceptibilities recorded for the penicillin antibiotics.

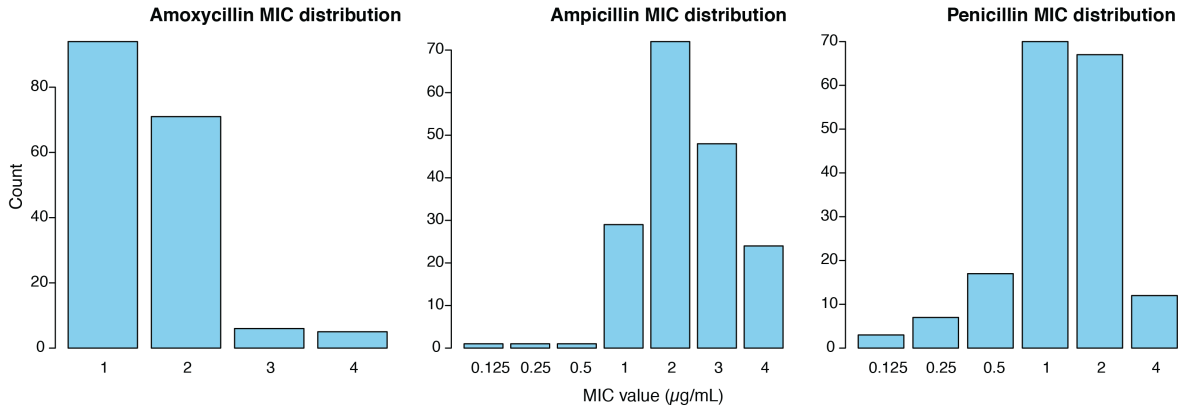


Figure 27: Counts of isolates within each MIC grade tested, for the penicillin class beta-lactams.

Most isolates exhibited reductions in susceptibilities to the penicillins, although no-extremely high levels of penicillin non-susceptibility were found. Susceptibilities however were generally at the intermediate to resistant levels according to CLSI/EUCAST guidelines.

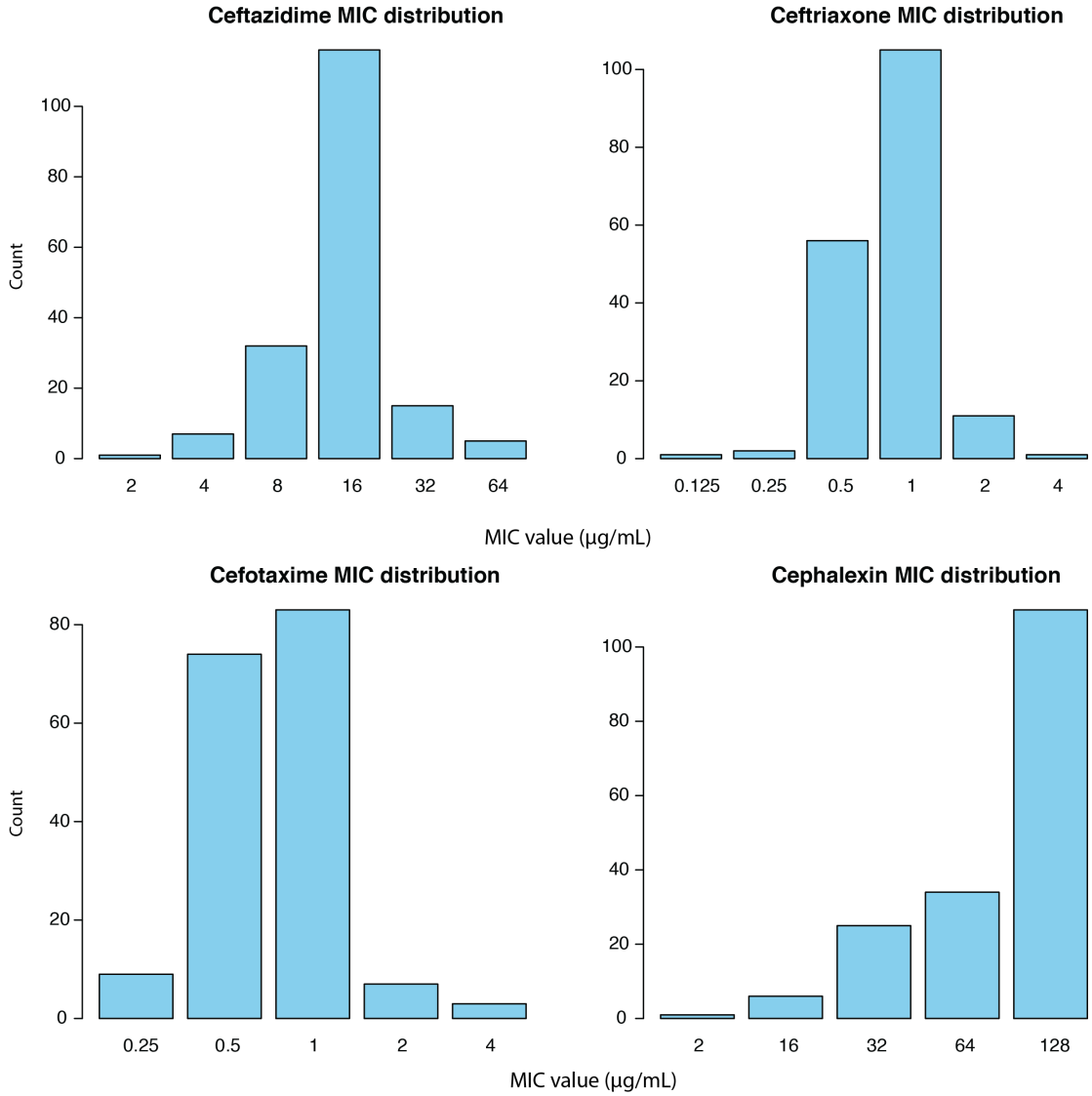


Figure 28: Counts of isolates within each MIC grade tested, for the cephalosporin class beta-lactams.

Cephalosporin resistance was similarly at the intermediate to resistant level for most PMEN1 isolates tested (Figure 28). For cephalexin and ceftazidime however, PMEN1 isolates showed very high levels of non-susceptibility; for cephalexin no upper limit for resistance could be identified.

3.5.2 Glycopeptides

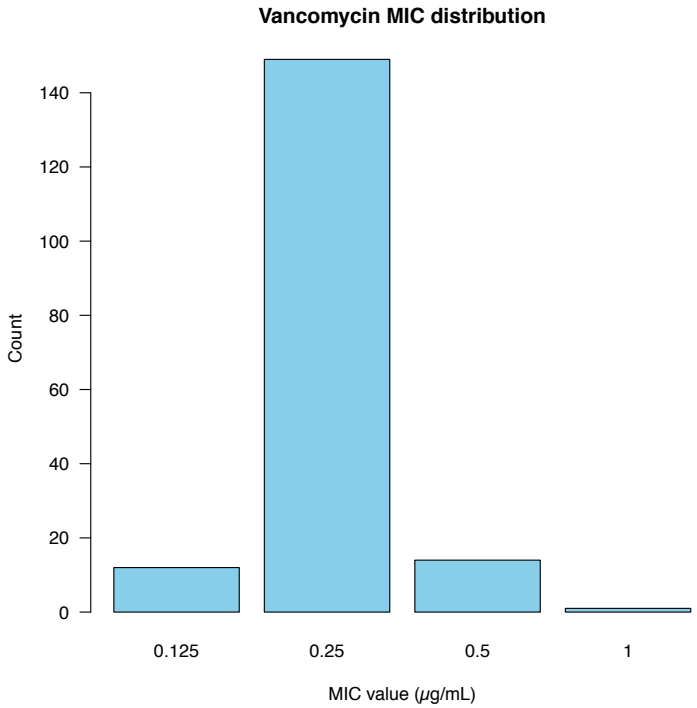


Figure 29: Counts of isolates within each MIC grade tested, for the only glycopeptide tested, vancomycin.

Susceptibilities to vancomycin, the only glycopeptide tested, were low, and non-standard MICs were recorded in order to better differentiate susceptibilities between PMEN1 isolates (Figure 29). Vancomycin remains one of the last effective anti-pneumococcal antibiotics (Novak et al., 1999). Widespread susceptibility to this antibiotic indicates that the success of the PMEN1 clone has largely been a result of resistance to the more commonly available beta-lactams. The PMEN1 lineage appears however to have begun to develop vancomycin resistance, which perhaps reflects an increasing reliance on this antimicrobial following the reduction in efficacy of other commonly used beta-lactams.

3.5.3 SNP variation and Antibiotic Susceptibility

To determine whether there was any association between a strains MIC value and the genetic variation present in the strains genome, a phylogenetic tree was constructed using RAxML. An MIC heatmap was then plotted alongside, to determine how beta-lactam susceptibility varied across the phylogeny.

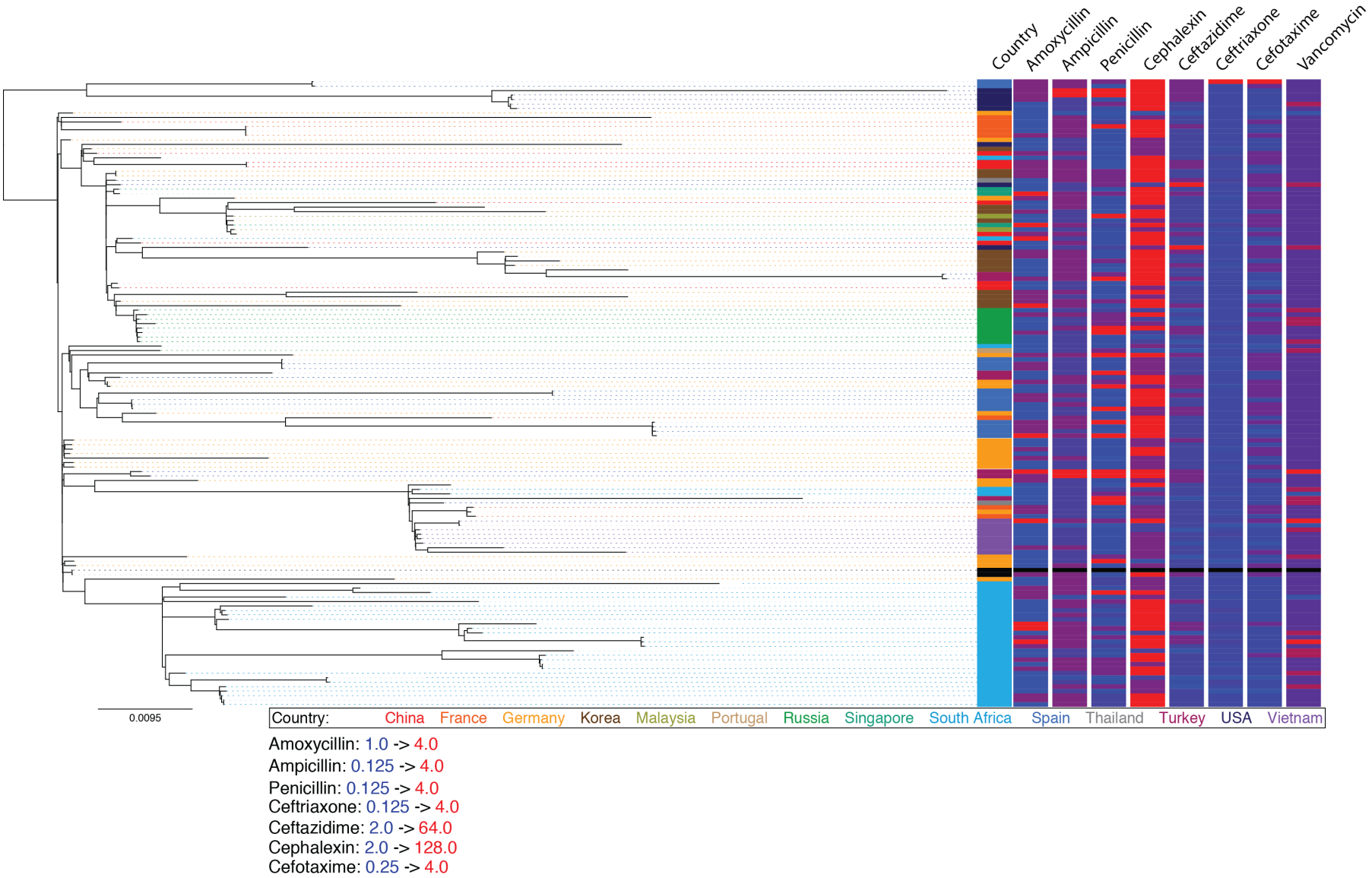


Figure 30: Phylogeny plotted based on genomic variation present among the PMEN1 isolates. Metadata for country of origin, and antibiotic susceptibility is plotted alongside. MIC data are plotted in the form of a heat map, from low MIC (blue) to high (red). Metadata fields for the ATCC700669 reference sequence used for phylogeny construction are coloured black.

Figure 30 was generally uninformative, with little clustering of resistant isolates to particular isolates on the phylogeny. Consequently, a more thorough investigation into *pbp* variation was undertaken. Here, a phylogenetic tree was constructed for each of the *pbps 1a*, *2b*, and *2x*, based on the non-synonymous nucleotide variation present. The corresponding phenotype information was then plotted alongside.

3.5.4 *Pbp2x*

Pbp2x was the most variable of the *pbp* genes, with 234 SNPs identified, consistent with mosaicism occurring within this gene. From this number, 72 SNPs were identified as being non-synonymous, and were used for phylogeny reconstruction (Figure 31). The phylogeny indicated that much of this variation was limited to three divergent branches, with very little tree structure identifiable. No relationship between the phylogenetic branching pattern and the country of origin was evident. Furthermore, MIC values were similarly found to vary across the phylogeny, apparently independent of the phylogeny.

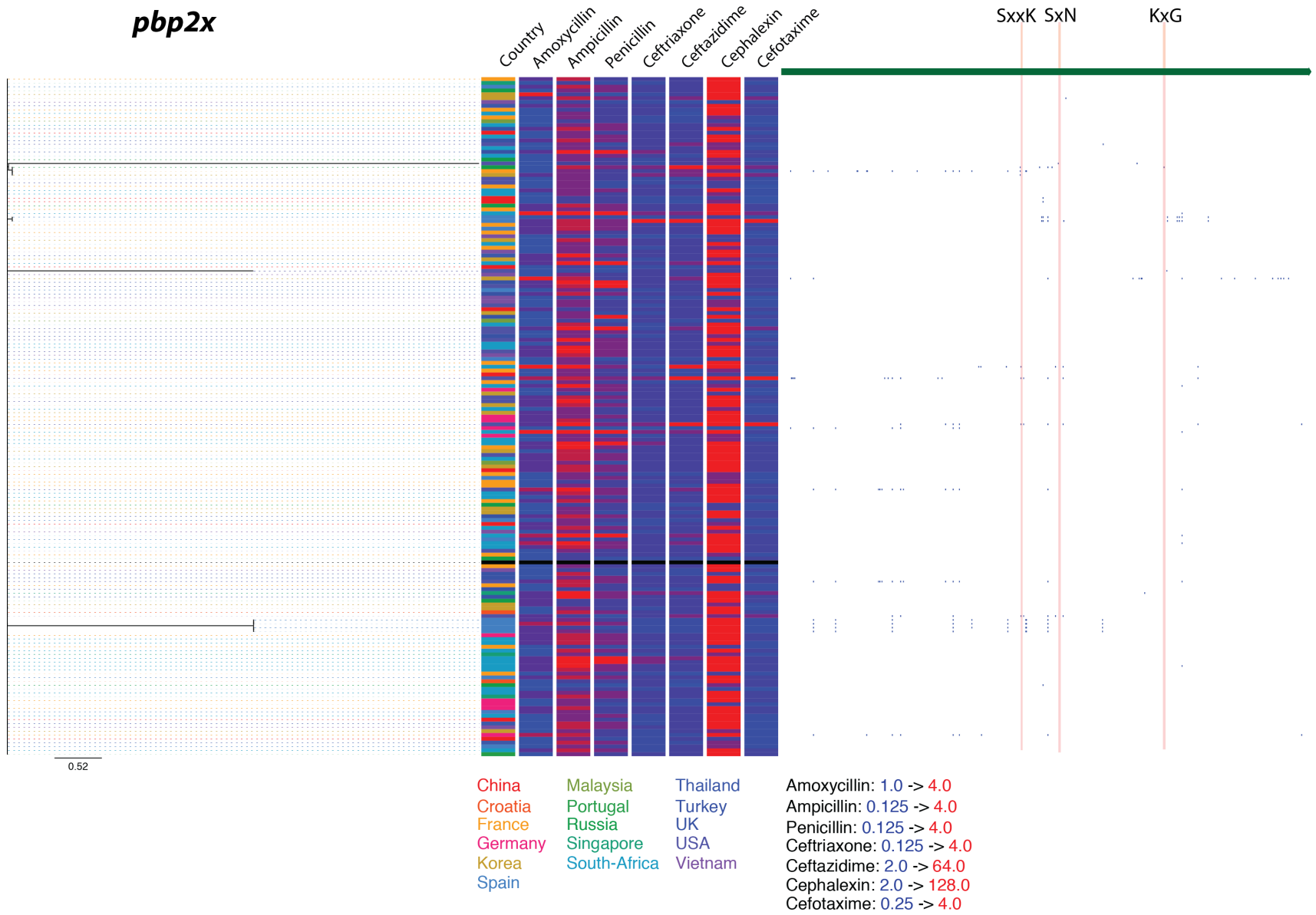


Figure 31: A phylogenetic tree based on non-synonymous nucleotide polymorphisms present in the *pbp2x* allele of PMEN1. The MIC value is plotted (low: blue, high: red) alongside, in addition to country of origin. The position of non-synonymous polymorphisms respective to the *pbp2x* gene, and the key motifs (SxxK, SxN, and KxG) is also plotted.

Table 12 also indicates, that in isolates where resistance associated SNPs, based on the literature, had reverted, there was no consistent increase in beta-lactam susceptibility. Such strains did generally carry a penicillin MIC value lower than the modal value for PMEN1 (2µg/mL), but was often not substantial reduced.

Isolate ID	Country of origin	Pen	Amp	Amox	Van	Ceftri	Cefta	Cepha	Cefo	Resistant AA in *pbp2x*	Type of change
4330_3_4	Germany	1	2	1	0.25	1	32	64	0.5	T338A	A->G
4311_8_5	South-Africa	2	4	4	0.25	0.25	64	128	0.5	Q281L	L->Q
4069_2_1	China	1	1	1	0.25	0.5	8	32	0.25	I371T	reversion
4232_8_2	Spain	0.5	1	1	0.125	0.5	16	32	0.5	I371T	reversion
4069_2_7	China	1	1	1	0.25	0.5	4	64	0.25	I371T	reversion
4232_8_12	Spain	2	3	2	0.5	4	64	128	4	I371T	reversion
4232_6_4	Russia	0.125	0.25	1	0.25	1	2	64	1	I371T	reversion
4232_6_2	Russia	2	3	1	0.25	2	64	128	2	R384G	G->S
4270_8_10	USA	0.25	1	2	0.25	1	32	64	2	S389L	reversion
4311_8_5	South-Africa	2	4	4	0.25	0.25	64	128	0.5	S389L	reversion
4021_5_9	Korea	1	3	4	0.5	1	32	128	1	N514H	reversion
4232_8_12	Spain	2	3	2	0.5	4	64	128	4	L565S	S->F
4232_8_2	Spain	0.5	1	1	0.125	0.5	16	32	0.5	L565S	S->F
4270_8_9	USA	2	4	2	0.25	2	64	128	4	T745K	reversion
4232_3_1	France	1	3	3	0.25	0.5	8	128	0.5	T745K	reversion

Table 12: Strains that have undergone reversal or additional changes to amino acid (AA) substitutions previously identified as associated with beta-lactam resistance in the literature. For context, MIC information is plotted, in addition to the resistance associated AA change, and whether this was a reversal or novel substitution.

3.5.5 *Pbp2b*

There was a similarly high level of genetic polymorphism present in the *pbp2b* gene, with 159 SNPs present, of which only 31 were identified as conferring non-synonymous changes (Figure 32). Similar to the *pbp2x* phylogeny, no clear relationship between the position of isolates on the tree and antimicrobial susceptibility was identified, with the phylogeny showing very little structure (Figure 32). A tight cluster of 6 strains isolated from South Africa in 2001 were apparent, being highly divergent from the rest of the tree. These isolates appeared similar in terms of their penicillin susceptibilities, and also shared similar genetic polymorphism. This variation was not limited to the motif regions discussed previously, but occurred throughout the gene. No other isolates were found to cluster in terms of country of origin, and MIC values were once again found to be highly mixed across the tree. No isolates were found to have undergone further modification, or reversal of the polymorphisms that had been identified previously as associating with penicillin resistance.

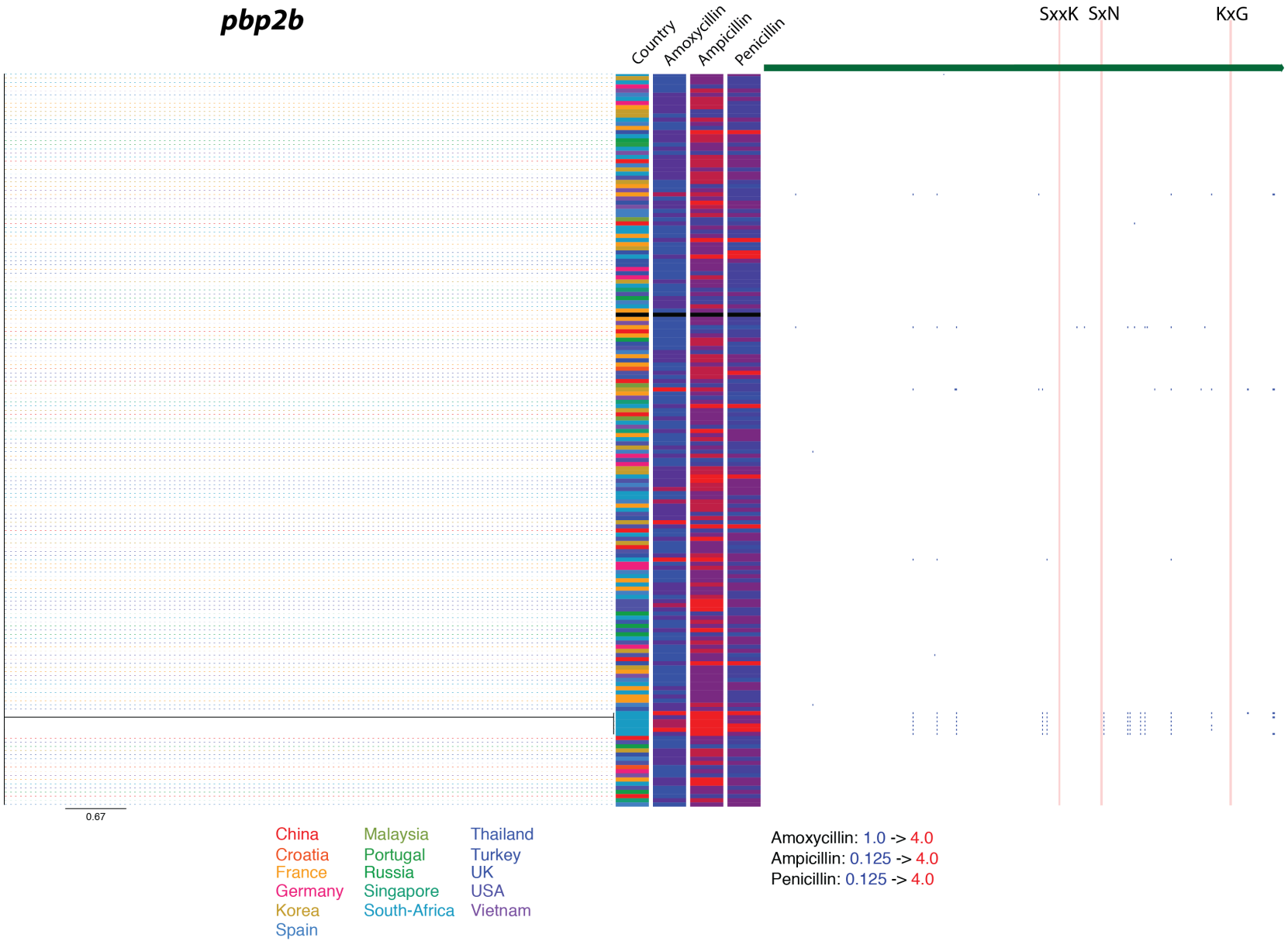


Figure 32: A phylogenetic tree based on non-synonymous nucleotide polymorphisms present in the *pbp2b* allele of PMEN1. The MIC value is plotted (low: blue, high: red) alongside, in addition to country of origin. The position of non-synonymous polymorphisms respective to the *pbp2b* gene, and the key motifs (SxxK, SxN, and KxG) is also plotted.

3.5.6 *Pbp1a*

The *pbp1a* locus was found to be far more conserved than either pbp2x or pbp2b, with 36 SNPs identified in total, of which 14 were identified as being non-synonymous. Interestingly, although variation was lower, at 38%, a greater proportion of this variation was non-synonymous, compared to *pbp2x* (31%), and *pbp2b* (19%). As *pbp1a* modifications are necessary for higher-level beta-lactam resistance (Smith and Klugman, 1998), this may indicate variation in *pbp1a* is less tolerated compared to *pbp's 2x* and *2b*. In addition, much of this variation appeared to be confined to just two divergent isolates, which had originated in South Africa, and France. Furthermore, the French isolate represented the non-PMEN1 strain BM4200, isolated prior to the spread of PMEN1. The association between these isolates on the phylogeny appeared to reflect a limited number of shared polymorphisms between these isolates. Neither strain was associated with high MIC values for any antibiotic. This could in part have been a reflection of their divergence from the rest of the PMEN1 lineage (Figure 33).

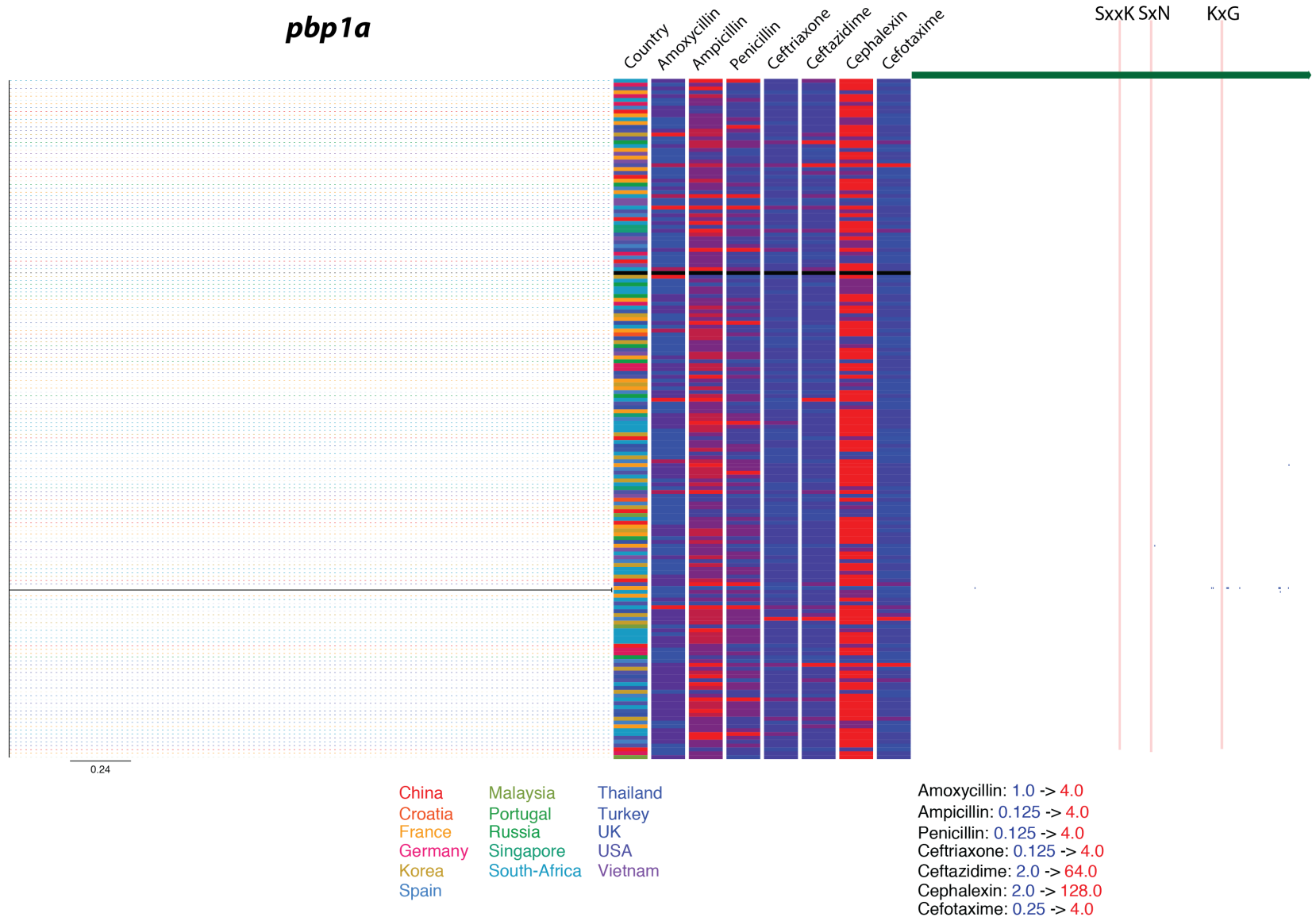


Figure 33: A phylogenetic tree based on non-synonymous nucleotide polymorphisms present in the *pbp1a* allele of PMEN1. The MIC value is plotted (low: blue, high: red) alongside, in addition to country of origin. The position of non-synonymous polymorphisms respective to the *pbp1a* gene, and the key motifs (SxxK, SxN, and KxG) is also plotted.

Given the high level of conservation across this gene, very few inferences could be made about the variation underpinning the different MIC values across the PMEN1 lineage. This suggested that comparing *pbp* alleles in isolation offered very little insight into the mechanisms of beta-lactam resistance underpinning MDR in the PMEN1 lineage. A method was therefore needed in order to consider the interactions between *pbp* alleles, and the wider genomic context.

3.5.7 Vancomycin

The mechanism of resistance to vancomycin involves genes outside of the cell wall synthesis pathway. Genes associated with the two component vncRS were therefore chosen for analysis. Although vancomycin resistance was not identified, by using non-standard MIC values, it was possible to identify variation in vancomycin susceptibility with the PMEN1 lineage. However, once again, no association between MIC value and nucleotide polymorphisms could be identified, with high levels of sequence conservation again present across the genes analysed (Figure 34).

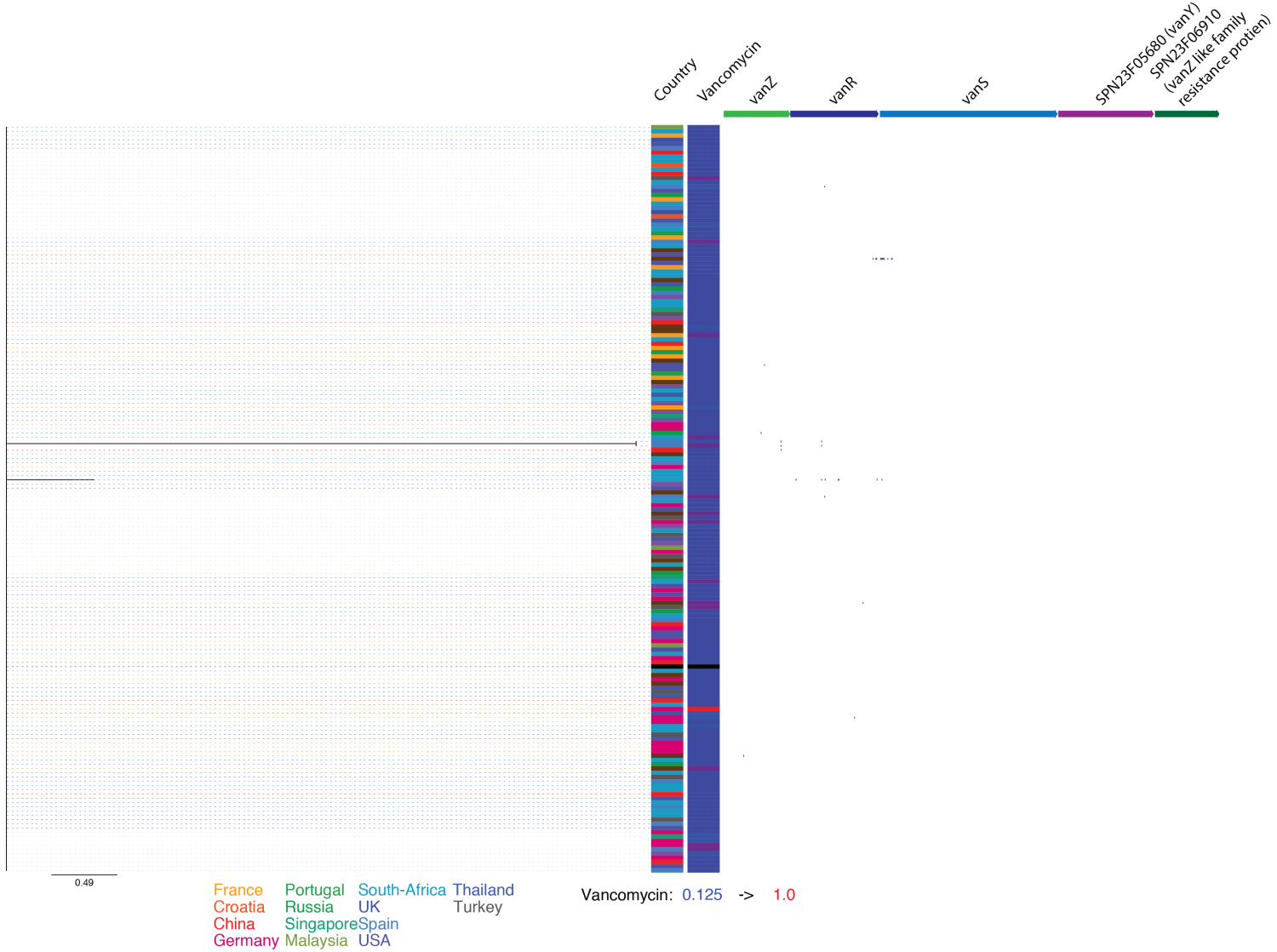


Figure 34: A phylogenetic tree based on non-synonymous nucleotide polymorphisms present in the genes associated the vncRS two-component system, responsible for vancomycin resistance in PMEN1. The MIC value is plotted (low: blue, high: red) alongside, in addition to country of origin. The position of non-synonymous polymorphisms respective to their positions in the genes used is also plotted.

3.6 GWAS methodology and Approach

3.6.1 Method development

The above analyses revealed that there was a general lack of relationship between the MIC information and the nucleotide polymorphisms present within the alleles considered. The interactions between *pbp* alleles are thought important for the development of beta-lactam resistance, and there is growing evidence to suggest other genes play a role in beta-lactam resistance. By inference, vancomycin resistance may also involve other mechanisms that were not considered here.

To investigate this further, a Genome-Wide Association Study (GWAS) was developed. The GWAS method was developed in the field of eukaryotic genomics as a way of statistically identifying associations between phenotype and genotype. Despite the widespread use and success of this approach, its use in the field of microbiology has been limited. This has largely been due to the limited availability of bacterial genomes for study, and the problems posed by clonality in bacteria. For prokaryotes, where recombination occurs infrequently, it can lead to high levels of conservation in gene order between isolates. Linkage disequilibrium, the term used to describe such non-random allele associations at different loci, can therefore lead to false associations occurring between genotypes and phenotypes (Slatkin, 2008, Chewapreecha et al., 2014b). Recombination can help to break up such associations, and consequently, sampling large numbers of isolates from diverse lineages can help breakup such non-random associations (Chewapreecha et al., 2014b). However, recombination can also cause issues for identifying phenotypically important effects, due to the occurrence of SNPs within the recombined fragment, which do not affect the phenotype of interest. Under conditions where the recombination event introduces SNPs that are positively selected, additional flanking SNPs may also associate with this phenotype, due to their location alone. Consequently, when studying MDR lineages such as PMEN1, such linkage affects needed to be considered when interpreting the results.

The current programs such as PLINK (Purcell et al., 2007) and EMMA (Zhou and Stephens, 2014) are also limited to the study of binary outcomes (Chewapreecha et al., 2014b, Gifford et al., 2013). In order to apply this methodology to the PMEN1 isolates, it was necessary to use a more flexible approach, as MIC values were recorded over a range of antibiotic concentrations.

To overcome this, a contingency table approach was developed in the framework of the Fisher's Exact test. This allowed for the construction of a contingency table using counts of SNP presence and absence against MIC values. A minor allele frequency exclusion criteria was used, excluding SNPs that were only present once in the dataset. Furthermore, to reduce unnecessary testing, only non-synonymous SNPs were considered in this study, or those present in intergenic regions, as these could represent promoter regions. Each SNP was then tested for its association with an antibiotic MIC, which was repeated for each antibiotic (excluding vancomycin). The p-values generated were then corrected for multiple testing using the q-value package in R (Dabney et al., 2004).

3.6.2 Cell Wall Synthesis Genes

To verify the approach, the program was first run using a subset of genes, known to be involved in the cell wall synthesis pathway (Table 13). All members of the PMEN1 lineage (excluding BM4200 strain) were tested. Given the high level of characterisation of the *pbp* alleles, the intention was to identify SNP associations within *pbp* alleles statistically, and subsequently confirm these associations based on the literature. By considering all genes known to be involved in the cell wall synthesis pathway, the then verified approach could be used to identify additional gene candidates involved in beta-lactam resistance.

Common name	Functional name	Promoter identified
pbp2x	penicillin-binding protein 2x	True
mraY	phospho-N-acetylmuramoyl-pentapeptide-transferase	False
pbp1a	penicillin-binding protein 1a	True
bacA	UDP pyrophosphate phosphatase	False
murM	serine/alanine adding enzyme	True
murN	alanine-adding enzyme	False
murD	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase	True
murG	UDP diphospho-muramoyl pentapeptide beta-N acetylglucosaminyl transferase	False
pbp3	D-alanyl-D-alanine carboxypeptidase	False
murZ	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	True
murB	UDP-N-acetylenolpyruvoylglucosamine reductase	True
murC	UDP-N-acetylmuramate--L-alanine ligase	False
murE1	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-L-lysine ligase	True
murE2	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase	False
murF	UDP-N-acetylmuramoyl-tripeptide--D-alanyl D-alanine ligase	False
ddl	D-alanyl alanine synthetase A	True
pbp2b	penicillin-binding protein 2b	True
murA1	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	False
pbp2a	penicillin-binding protein 2a	True
pbp1b	penicillin-binding protein 1b	True

Table 13: Genes present in the cell wall synthesis pathway that were used in the first analysis. Where possible the promoter for each gene was identified using the tool PEPPER (pepper.molgenrug.nl).

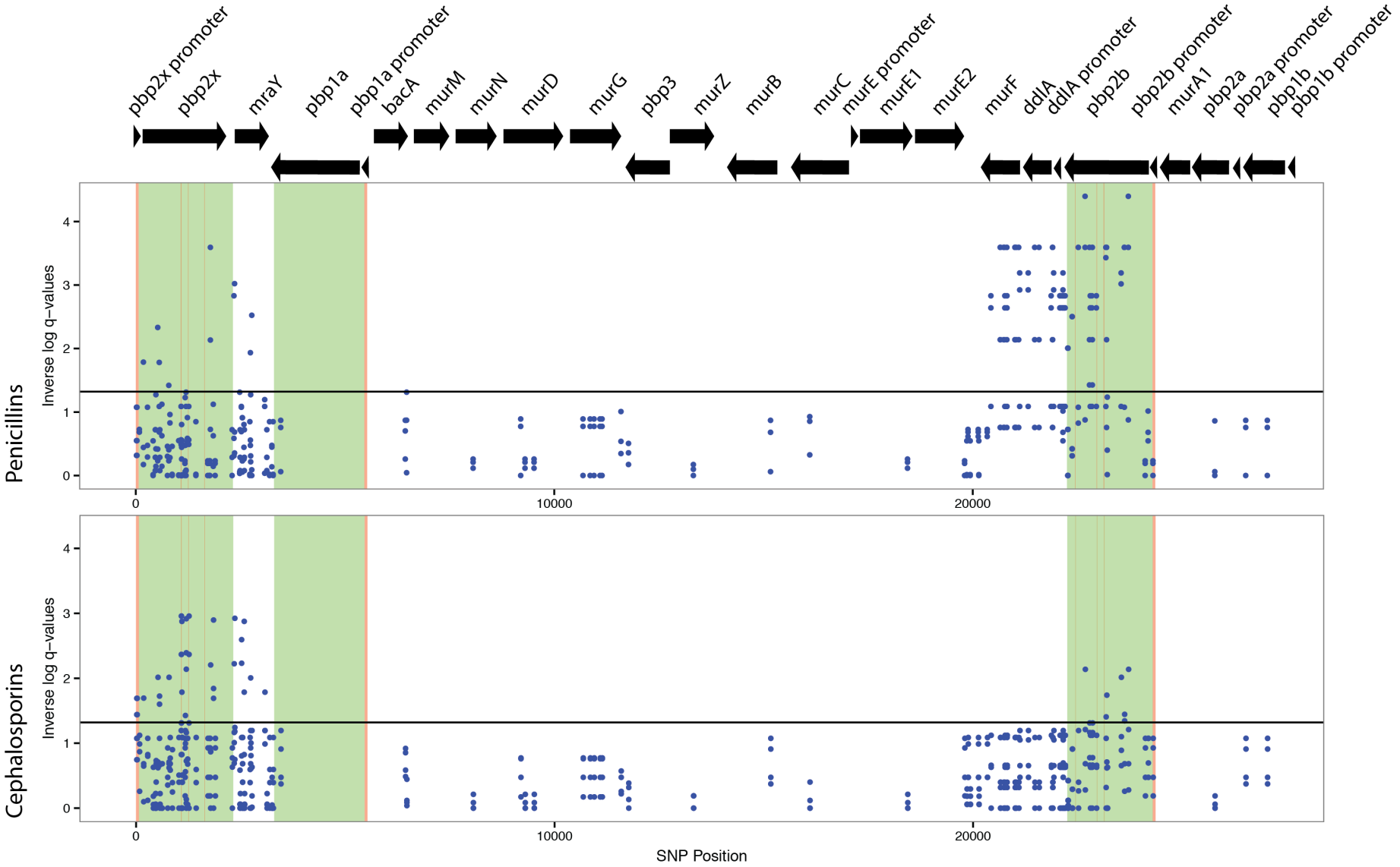


Figure 35: Manhattan plot for SNP associations restricted to the cell wall synthesis genes. SNP positions relative to the genes in the peptidoglycan synthesis pathway are plotted, with genes indicated at the top. The grey horizontal lines indicate the cut-off q-values, with significance increasing above this line.

Figure 35 shows the associations recorded, and the q-value cutoff value used (q= 0.5). The approach successfully identified variation within *pbp2b* and *2x* alleles as associated with beta-lactam resistance. No associations were found *pbp1a*, which was perhaps in part due to the limited variation present in this gene. Several non-*pbp* genes were identified as conferring penicillin resistance, however each of these genes was located close to either *pbp2x* or *pbp2b* in the host genome. Furthermore, cephalosporin resistance is not thought to involve *pbp2b*, yet a limited number of SNPs did associate with resistance at this locus. Such associations could have resulted from co-selection for resistance to the penicillins in addition to cephalosporin's, with a similar affect having been identified previously (Chewapreecha et al., 2014b).

Gene	Amoxycillin	Ampicillin	Penicillin	Ceftazidime	Cefotaxime	Ceftriaxone	Total Resistance associated SNPs
ddlA	10	10					10
ddlA promoter	8	8					8
mraY	5			5	2	1	8
murF	10	10					10
murM promoter			1				1
pbp2b	18	11		8		1	20
Pbp2x	3	1		8	8	5	10
pbp2x promoter				1		1	1

Table 14: Genes containing SNPs that were identified as being significantly associated with a loss in beta-lactam susceptibility. For each antibiotic tested, the number of SNPs within each gene that significantly associated with resistance to that antibiotic is indicated. Total counts of resistance associated SNPs in each gene are shown in the final column, indicating that some SNPs were found to correlate with resistance to more than one antibiotic.

Table 14 indicates the genes found to carry variation that significantly associated with a change in beta-lactam MIC. Variation in the promoter regions identified for *ddlA*, *murM* and *pbp2x* were all found to potentially alter the MIC values recorded. *Pbp2b* carried the greatest number of resistance associated SNPs, many of which appeared to confer changes in susceptibilities to penicillins and cephalosporins as discussed above. High number of resistance associated SNPs were also identified in the genes immediately flanking *pbp2b- ddlA* and *murF*.

Figure 36 indicates the number of SNPs that associated with a loss of beta-lactam susceptibility to one or more antibiotic. Two classes of beta-lactams, the penicillins and the cephalosporins had been included in this study. Consequently, SNPs may be expected to confer a loss of susceptibility more often to other members of the same class. Despite this, in several cases, Figure 36 indicates that SNPs significantly associated with a loss in beta-lactam across antibiotic classes, such as in the case of 2 SNPs in *mraY*, as well as *pbp2x,* and in *pbp2b*. Such associations could result for co-selection under the pressure of different associations, as suggested above (Chewapreecha et al., 2014b).

Figure 36 also indicates that SNPs commonly conferred a loss in susceptibility to amoxicillin and ampicillin, suggesting similar target modifications are required for resistance to arise to these antimicrobials. Whilst a similar association between cefotaxime and ceftazidime was apparent, such associations often appeared to correspond to a very limited number of SNPs, suggesting a weaker relationship. The analysis was generally uninformative however for penicillin susceptibility, with only a single SNP in a predicted murM promoter region being found to associate with a change in penicillin MIC.

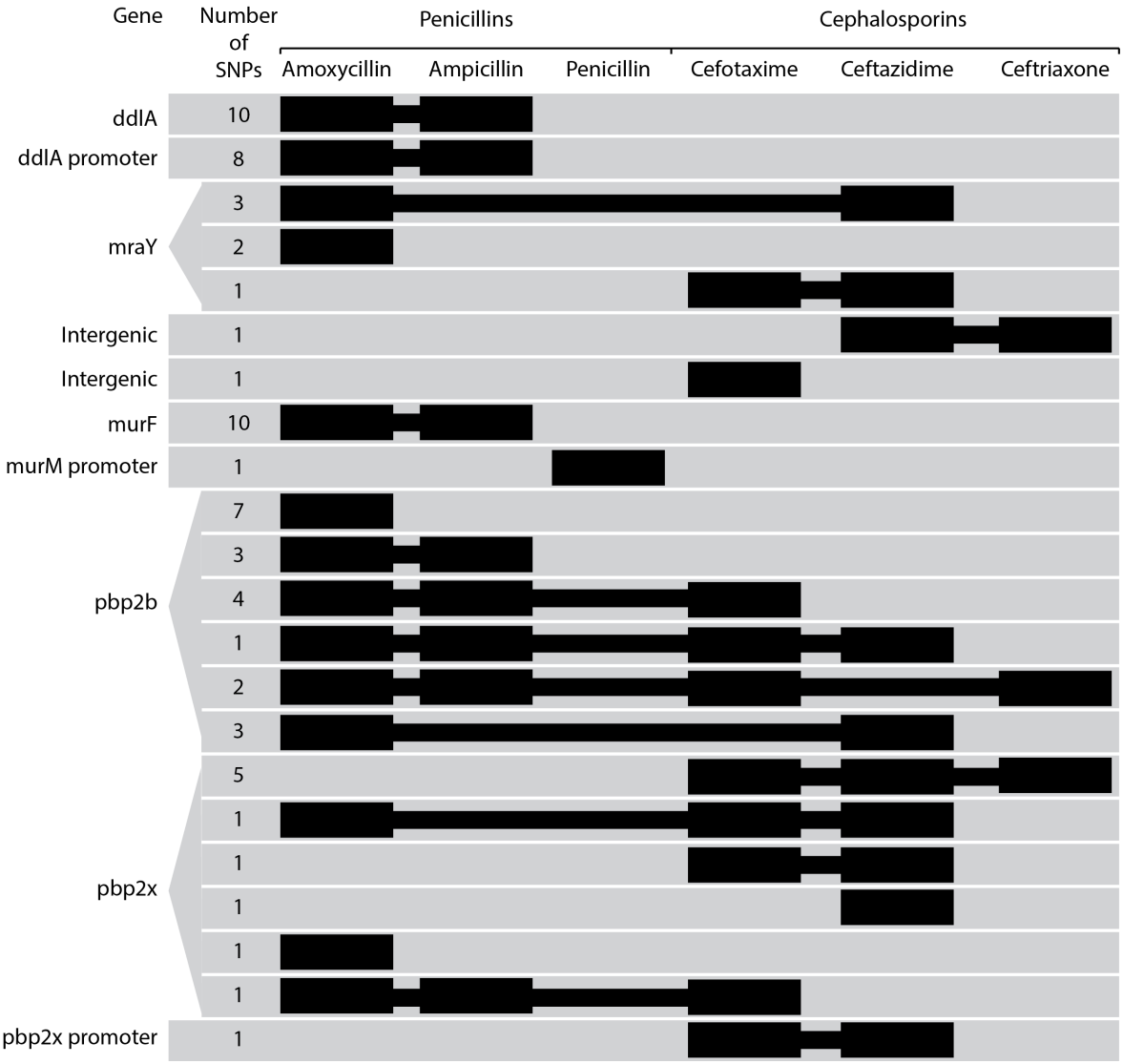


Figure 36: For each gene, the numbers of SNPs associated with susceptibility changes to one or more antibiotic is indicated. For example, in *ddlA*, 10 SNPs significantly associated with a loss in susceptibility to amoxicillin in addition to a loss in susceptibility to ampicillin.

3.6.3 Genome-wide expansion

To determine whether additional associations could be identified, the analysis was repeated using genome-wide nucleotide polymorphisms, rather than being restricted to cell wall synthesis genes. In total, this included 18806 SNPs, either being identified as non-synonymous, or occurring intergenically.

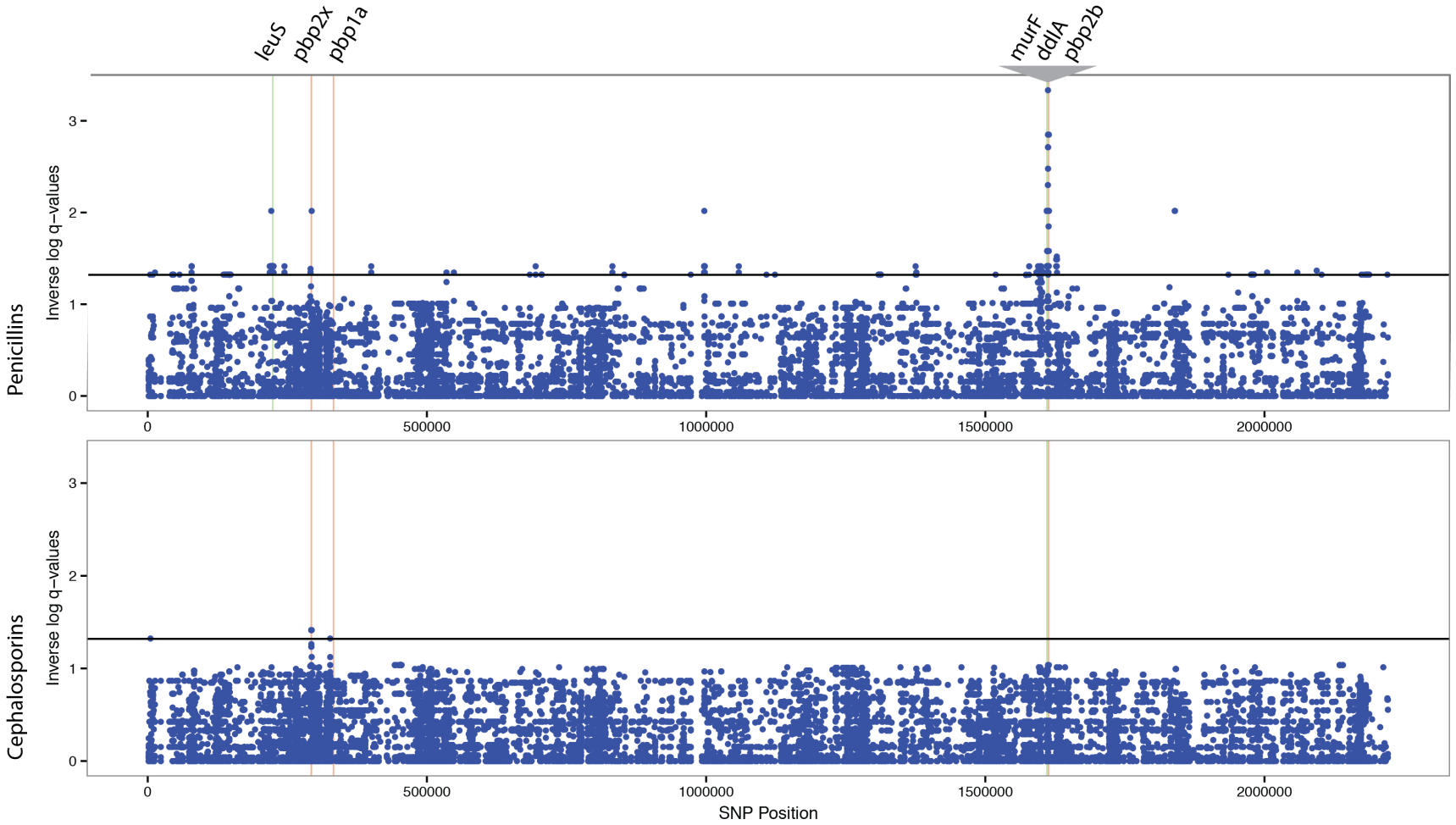


Figure 37: Manhattan plot of SNPs associated with changes in beta-lactam susceptibilities performed genome-wide. Key associations are indicated, in addition to the cutoff value over which significance occurs.

Figure 37 shows the Manhattan plot for genome-wide polymorphisms as they associated with MIC changes (q=0.05), and a full list of the genes containing variation statistically associated with changes in beta-lactam susceptibility is given in appendix 10.9.

From the Manhattan plot (Figure 37), although a number of SNPs were identified as associating with resistance, few of these clustered into defined peaks, which is expected to occur in true SNP-phenotype associations. Therefore, it is likely that many of the positive associations identified are an artefact of recombination and the hitchhiking of SNPs on recombined sequences, shared between isolates. This is particularly likely in cases where the recorded significance was also low. The Manhattan plot is once again divided into two, to reflect SNPs that were associated with changes in susceptibility to penicillins and those associated with changes in cephalosporin susceptibilities. In both plots the location of *pbp* genes are labelled and indicated with red lines. In the penicillin plot, a clearly defined peak occurred in the region containing the *pbp2b* allele. Significant associations were also identified for SNPs in the *pbp2x* gene, although, these associations were far fewer and weaker.

To hone in on the variation present in *pbp2x* and *pbp2b* genes, Figure 38 indicates the levels of SNP associations present in these genes, coloured according to the antimicrobial being tested. This indicates that many of the significant associations were once again limited to amoxicillin and ampicillin within this class. Furthermore, little of this variation was identified as occurring near the conserved motif regions indicated in the figure. Such associations could therefore more likely reflect homology, rather than truly affecting antimicrobial susceptibilities.

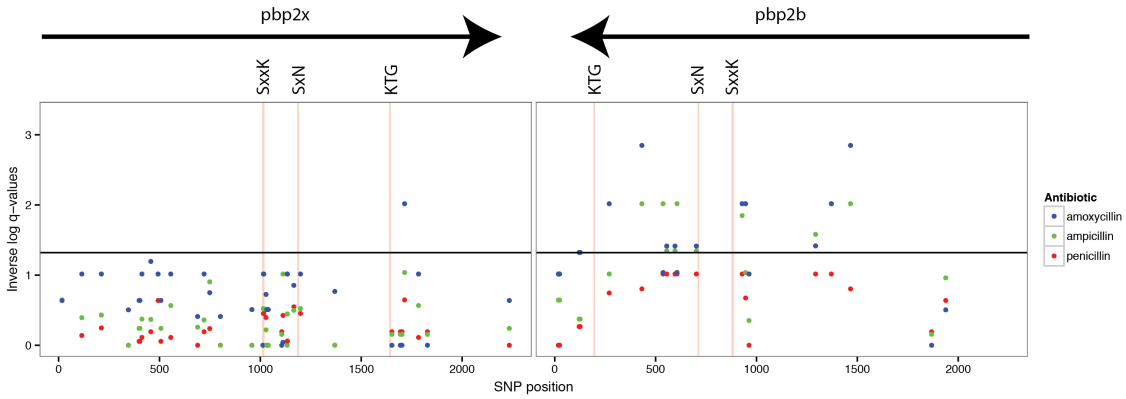


Figure 38: Indicated for the penicillin class beta-lactams are the SNP associations identified, coloured according to the antibiotic being tested. Here the strongest associations can be found to occur for amoxycillin.

Outside of the *pbp* genes, a total 21 SNPs present in the leucyl-tRNA synthetase gene (*leuS*) were found to associate with resistance with a change in susceptibility to amoxicillin and ampicillin. This gene has not previously been found to associate with a resistance phenotype in pneumococci. The *leuS* gene is also located 66701bp upstream of the *pbp2x* gene, a distance substantial enough to suggest that recombination events affecting the *pbp2x* gene would be unlikely to extend as far as *leuS* (Croucher et al., 2012). As such, this suggest that the association of SNPs within the *leuS* gene is not due to a hitch-hiking effect. Additional SNP associations were also found in the genes *fsaA* and *gldA,* corresponding with changes in amoxicillin and ampicillin susceptibilities. However, both genes are located adjacently to the *leuS* gene. Furthermore, the associations of *fsA* and *gldA* were weaker, and fewer- a total of 6 and 4 SNPs associated with susceptibility changes in *fsA* and *gldA* respectively, compared to 21 in *leuS* (Table 15). As such, this suggests the associations within flanking *fsA* and *gldA* likely occurred due to linkage across this region due to recombination primarily affecting the *leuS* locus. Studies in *E. coli* found that resistance to amdinocillin, a beta-lactam, was moderated by *leuS* activity through it's effect on the guanisine-5'-diphosphate 3'-disphosphate (ppGpp) pool. PpGpp availability moderates the inauguration of the "stringent" response in *E. coli*, a stress response, which is characterised by the redirecting of resources away from cell division and growth and towards amino-acid synthesis, aiding survival (Jain et al., 2006, Ross et al., 2013), suggesting a possible mechanism by which leuS

modification could affect beta-lactam susceptibility in other species- by affecting the sequestering of PBP proteins during cell growth and division.

Gene	Amoxycillin	Ampicillin	Cefotaxime	Ceftazidime	Ceftriaxone	Penicillin	Unique
ddlA	10	4					10
ddlA promoter	8	8					8
fsaA	4	6					6
gldA	3	4					4
ileS						9	9
leuS	21	21					21
murF	11	4					11
nanB		3				3	3
pbp2b	12	10					14
pbp2x	1		2	4			7
recR	5	2					5

Table 15: Genes containing SNPs that were identified as being significantly associated with a loss in beta-lactam susceptibility. For each antibiotic tested, the number of SNPs within each gene that significantly associated with resistance to that antibiotic is indicated. Total counts of resistance associated SNPs in each gene are shown in the final column, indicating that some SNPs were found to correlate with resistance to more than one antibiotic.

Significant SNP associations were also identified in the genes flanking *pbp2b*. In particular this included *murF* and *ddlA*. The association between variation in *ddlA* and *murF* and changes in beta-lactam susceptibility has been identified previously. This association has previously been thought to reflect linkage, due to recombination affecting the *pbp2b* allele (Enright and Spratt, 1999). This view is further supported by association of SNPs occurring in the similarly *pbp2b* flanking genes *nanB*, *recR*, and *ileS,* yet these genes have a much more tenuous link to beta-lactam target proteins (Iaccarino and Berg, 1971, Berry et al., 1996, Sauerbier et al., 2012b). Therefore, although associations of *nanB*, *recR* and *iles* were weaker, and fewer in number, (Table 15) the association of *murF* and *ddlA* with changes in beta-lactam susceptibility most probably reflects linkage disequilibrium, due to selection focussed on *pbp2b*. The involvement of DdlA and MurF in cell wall synthesis, does however suggests possible mechanism by which they could affect beta-lactam susceptibility. It therefore remains to be tested *in vitro* whether recombination affecting the *pbp2b* locus also affects the activity of these adjacent genes.

Far fewer significant associations were identified when testing for SNP associations with cephalosporins (Figure 37). Here, a weak association was identified for a limited number of SNPs occurring in the *pbp2x* gene alone (Table 15). Whereas previously, a number of SNPs had been found to associate with resistance two both penicillin and cephalosporin classes, this association was also lost on expanding the analysis genome-wide. This further suggested co-selection and linkage had a stronger affect on the initial analysis, which was diluted when considering nucleotide variation genome-wide.

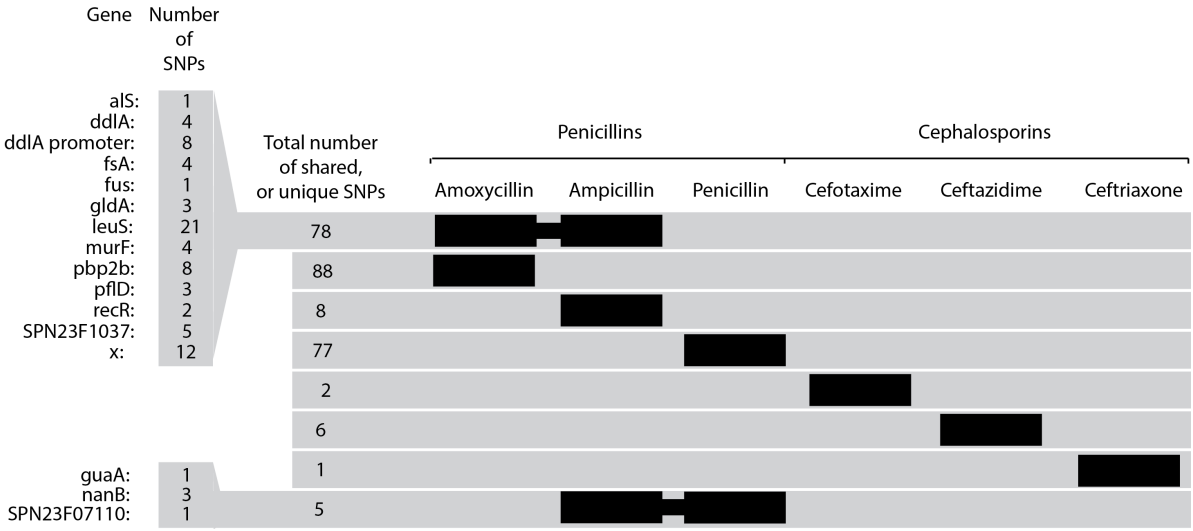


Figure 39: The numbers of SNPs associated with susceptibility changes to one or more antibiotic is indicated. For example, in total 78 SNPs significantly associated with a loss in susceptibility to amoxicillin in addition to a loss in susceptibility to ampicillin. Space limitations prevent the occurrence of SNPs being plotted for all genes (see appendix 10.9 for full list).

A similar plot to that in Figure 36 was constructed to visualise how SNP associations correspond to changes in beta-lactam susceptibility to one or more antibiotic (Figure 39). Here, it can be seen that most of the significant associations identified corresponded to changes in amoxycillin susceptibility alone. The relationship between amoxicillin and ampicillin however did remain, with 78 unique SNPs associated to MIC changes to both of these antimicrobials. Although 77 SNPs were found to associate with changes in penicillin susceptibility, such associations occurred randomly across many genes, rather than clustered within a few genes. Consequently, such associations are more likely to result from random linkage, rather than truly affecting susceptibility to this antimicrobial. One reason that stronger associations may not have been identified for penicillin may be related to the finding that the reference isolate, ATCC700669, contained large numbers of SNPs already identified as affecting penicillin susceptibility. Furthermore many of these were highly conserved across the PMEN1 lineage.

3.6.4 Vancomycin Resistance

No SNP associations with changes in vancomycin susceptibility were identified during the genome-wide study (Figure 40). The position of *pbp* alleles was plotted for reference, and in order to identify any linkage affects due to selection

at these sites. The lack of any SNP associations is perhaps due to the limited variability in vancomycin susceptibility across the PMEN1 isolates, so that only subtle changes are present within the lineage, which were not identified using this approach.

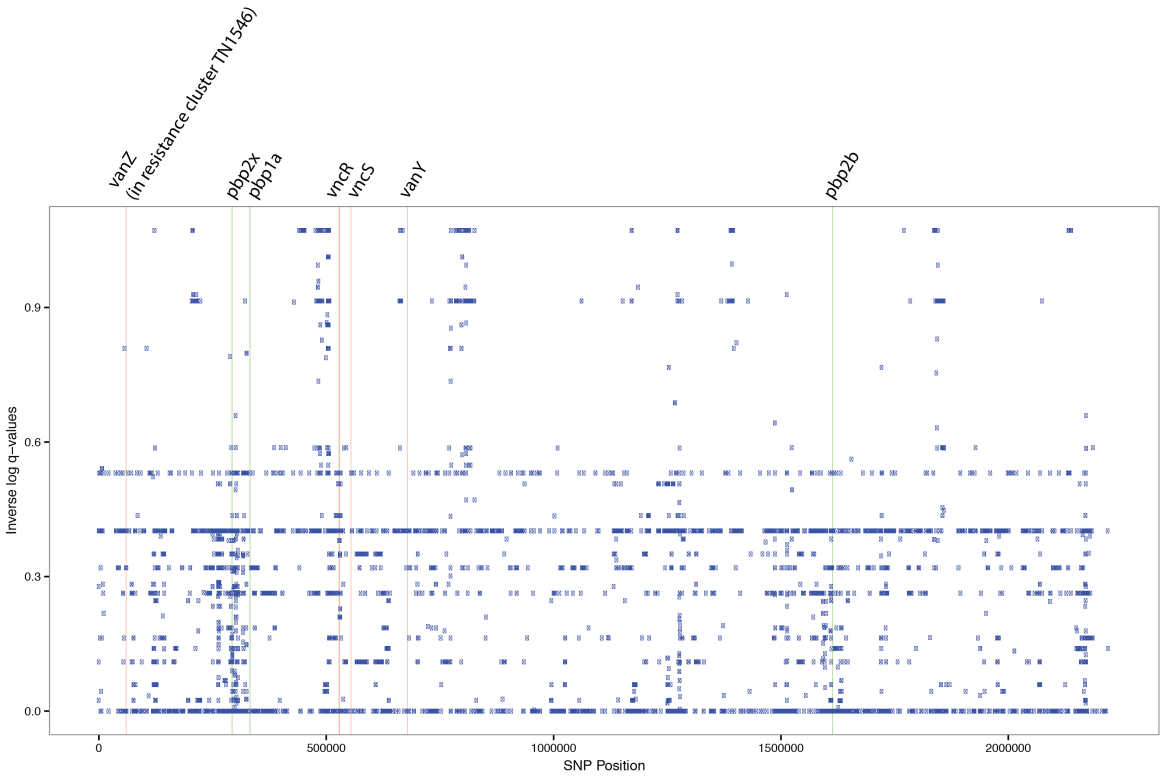


Figure 40: Manhattan plot for the GWAS analysis using vancomycin. No significant associations were identified. The position of known vancomycin resistance associated genes is shown. In addition the positions of *pbp* genes are labelled for context.

3.7 Discussion

This study represents one of the first attempts to use the GWAS approach to identify genetic variation important for the development of antimicrobial resistance within bacterial species. Whilst the application of GWAS to bacterial datasets has been limited to date due to concern about the occurrence of false positives, it has been demonstrated here that the approach can successfully identify genes known to be involved in beta-lactam resistance. Recombination can remain problematic to such analyses, due to the hitchhiking of SNPs within such fragments, which has to be considered when interpreting the results from such association studies. In this respect, the high level of characterisation available for the detection of recombination events (Croucher et al., 2015), in

conjunction to the availability of high quality genome annotations can help identify candidate resistance genes for further *in vitro* testing. The ease with which bacteria render themselves to laboratory experimentation is therefore likely to facilitate the wider use of GWAS in prokaryotic research- allowing clinical resistance targets to be identified, and verified *in vitro*.

3.7.1 Antibiotic Susceptibility in a globally important lineage

Whilst PMEN1 was not found to have majorly high levels of antimicrobial resistance, it had reduced susceptibility to all beta-lactams tested, and vancomycin susceptibility was altered in some isolates. PMEN1 originated under a regime of heavy beta-lactam usage in Spain, however its success appears to have been achieved without the necessity for very high levels of beta-lactam resistance. This is further corroborated by the finding that resistance to other classes of antimicrobial have arisen several times within the lineage, in addition to capsule switching events, which have allowed PMEN1 to subjugate vaccine pressures. Its success is therefore thought to largely reflect its ability to rapidly respond to clinical interventions (Croucher et al., 2011).

The MIC results recorded here indicate that reduced susceptibility to beta-lactams has however been relatively conserved, independent of country of isolation, or age of isolation, consistent with beta-lactam resistance not having affected the fitness of PMEN1 isolates relative to competing strains (Orio et al., 2011). Given the universal vancomycin susceptibility it appears the success of this clone is in part due to the limited availability of this antimicrobial. However, vancomycin MICs were not consistent across the lineage. Therefore, it is likely that PMEN1 isolates would have soon developed resistance to this antimicrobial if the selection pressure had been greater. Consequently, serotype switching of PMEN1 isolates, coupled with an ability to rapidly develop MDR means this lineage still represents in important threat to pneumococcal treatment globally. Interestingly, in several scenarios PMEN1 has began to decline, but this is mainly due to the spread of other PMEN lineages, suggesting an important role for

competition in the spread of such lineages (Croucher et al., 2013, Chewapreecha et al., 2014a).

3.7.2 The GWAS approach and Comparison with other studies

The current methods of choice for GWAS studies include PLINK and EMMA. Both, of which were developed for human genome comparisons, and are favoured for comparison of binary variables, i.e. resistant versus susceptible. Consequently, the method developed here was intended to draw more meaningful conclusions as to the SNPs that lead to the alteration of resistance, in an already MDR pneumococcal clone. The approach used here successfully identified the association between *pbp* modifications and beta-lactam susceptibility changes. It is also clear that linkage disequilibrium between genes was an important consideration when interpreting the results. To further this study, the role that recombination has had in mobilising resistance associated SNPs could be investigated further, similar to chapter 5.

A similar attempt to identify SNPs associated with beta-lactam resistance was carried out using independent genome datasets from pneumococci isolated in Thailand and Massachusetts (Chewapreecha et al., 2014b). This analysis was carried out using the tool PLINK to analyse nucleotide and indel diversity. This analysis used 3,085 pneumococcal genomes isolated from carriage in Maela, Thailand, (Chewapreecha et al., 2014b) and a second dataset which was composed of 616 carriage samples isolated from Massachusetts (Croucher et al., 2013). The authors of this study aimed to use large sample numbers to overcome linkage disequilibrium.

Whilst increasing the sample size should reduce the number of non-random associations between alleles resulting from linkage disequilibrium, genes flanking the *pbp* loci, and unrelated to cell wall synthesis were still identified as significantly associating with changes in beta-lactam susceptibility. Chewapreecha and colleagues (2014b) did not confirm the effects identified *in vitro* and consequently the degree to which linkage disequilibium was removed

cannot be assessed. Interestingly there was also relatively little overlap between the resistance-associated SNPs identified in the Massachussetts isolates compared to those detected in the Maela set (15% shared). This result is consistent with the general consensus, that beta-lactam resistance can arise by many different mechanisms. Chewapreecha and colleagues (2014b) were also limited in terms of the numbers of beta-lactams tested, with MIC information available for penicillin and ceftriaxone alone. In contrast, the analysis described here was carried out using MIC values recorded over 7 different antimicrobials. Providing an opportunity to determine how SNPs associated with a change in susceptibility to different classes of beta-lactam, or within a class.

Furthermore, by limiting this analysis to a globally successful MDR lineage it was hoped that this would offer greater insight into those pathways leading to a loss in beta-lactam susceptibility, without being harmful to the hosts fitness.

Comparison with Chewapreecha and colleagues' study (2014b) however suggests that although *pbp* gene modification is the driver of beta-lactam resistance, the modification of genes that can affect cell division, such as *leuS* in this study, and *FtsL* (Chewapreecha et al., 2014b) seem to affect resistance, but this has not been tested *in vitro*. Furthermore, given the limited overlap between the associated resistance genes in studies so far, whilst a general indication of the genes that can facilitate beta-lactam resistance may become evident, it appears that multiple pathways to resistance may be circulating in *in situ* as well as was found through *in vitro* study.

3.7.3 The problem of recombination and species divergence

Developing robust statistical methods for determining the associations between SNPs and phenotypes remains problematic. Although a Fisher's Exact test was used here, many false positive associations appear to have been identified. This problem comes from relatedness between strains, and homology between recombined fragments of genetic material between strains. The approach used here was simplistic, and did not attempt to correct for such non-random

associations. To overcome this, prior knowledge of gene function and recombination across the dataset was used to inform and interpret the gene associations identified. Whilst this is subjective, *in vitro* experimentation of candidate genes identified *in silico* could overcome this drawback. The ease with which bacteria can be grown and experimented on *in vitro* should aid the utility of the GWAS approach despite the complications resulting from recombination. Furthermore, there is currently no method available to remove the effects of recombination when identifying associations in this way- particularly when resistance is associated with gene mosaicism, as in resistance to the beta-lactam antibiotics.

3.7.4 Genes identified and the effect of hitchhiking

The effect of hitchhiking is well demonstrated in the case of *pbp2b*. Here multiple genes around the *pbp2b* allele have been identified as carrying polymorphisms that associate with beta-lactam resistance. Interestingly, genes flanking *pbp2x* and *pbp2b* similarly associated with beta-lactam resistance in Chewapreecha and colleagues' study (2014b). In the analysis presented here, of the genes carrying polymorphisms that associated with changes in beta-lactam susceptibility only MurF and DdlA were found to have any direct relationship with beta-lactams through their involvement in the cell synthesis pathway. The association of *murF*, *ddlA* and *recR*, which occur most closely to *pbp2b*, could therefore reflect the frequent recombination that affects this locus. In addition, the involvement of *pbp2b* modification with the early stages of penicillin resistance means that selection should occur earlier in response to antibiotic pressure here, than for instance in *pbp1a* (Smith and Klugman, 1998). Genetic linkage due to widespread recombination across the *pbp2b* gene is therefore likely to occur. However, whether the variation introduced during such events also affects MurF and DdlA activity is currently untested *in vitro*.

3.7.5 Associations between Antibiotics

The initial study was conducted using the cell wall synthesis pathway genes alone, which indicated a number of SNPs corresponded with a loss in beta-

lactam resistance to more than one antibiotic, and between antibiotic classes. On expanding the analysis genome-wide however, none of these inter-class associations were found. This suggests that the original associations resulted largely from genetic linkage, and a tendency for more or less variable strains to carry similar MIC values to all antimicrobials tested likely due to co-selection for different antibiotics (Chewapreecha et al., 2014b). Such genetic linkages therefore appeared to have been broken down when expanding the analysis genome-wide. In contrast, the relationship between amoxycillin and ampicillin remained strong under both analyses, suggesting similar modes of resistance to both of these antibiotics exist in PMEN1. The lack of association with penicillin may reflect the fact that although these three antimicrobials belong to the same class, both amoxicillin and ampicillin are part of the second generation penicillins, and so will have a slightly different mode of action.

3.7.6 Study Limitations

Although far fewer samples were analysed in this study compared Chewapreecha and colleagues' (2014b), the greater limitation stems from the focussing on a single lineage. As such linkage affects were widely identified, which complicated the interpretation of subsequent results. The deliberate interest in a single pneumococcal lineages means that recombinogenic blocks of sequence are likely to be highly conserved across clusters of isolates, leading to a greater likelihood of false positive associations, which is particularly likely to occur in genes flanking *pbp2b*. Whilst this limitation could be overcome through increasing the sample diversity, there are also limitations imposed by the throughput of the MIC testing method used here. Furthermore, automated methods for MIC testing are currently costly, and are rarely carried out routinely. Whilst disk diffusion testing is easier and more widely carried out for routine diagnostics, MIC testing across a range of values is more informative, and can allow for more meaningful conclusions to be drawn on the emergence of antimicrobial resistance. As such, the ability to scale-up MIC testing of clinical isolates is an important inhibitor on the ability to carry out future large-scale comparative genomic analysis.

3.8 Conclusion

The method developed here shows that even a simplistic application of the GWAS approach, without correcting for interrelatedness between strains, and between shared recombinogenic blocks can successfully identify genes whose modification can result in a loss of susceptibility to beta-lactam antibiotics. Sample diversity can improve the resolution of these associations, with a greater number of associations occurring likely due to linkage being identified in this study compared to the analysis carried out by Chewapreecha and colleagues (2014b). A further aspect however is the ability to now study genomes rather than particular genes of interest (Dowson et al., 1989, Hakenbeck et al., 1999), which was similarly found to be an important factor in reducing the influence of non-random SNP associations, by breaking-up the effects of linkage during statistical analysis.

Whilst, due to a lack of sample diversity, linkage was found to have affected this analysis, by studying a smaller number of samples than Chewapreecha and collagues (Chewapreecha et al., 2014b), it was possible to gather MIC information over a greater number of antibiotics. This offered insight into the modes of resistance to these antimicrobials occurring in PMEN1- suggesting a similar mechanism of resistance existed for both amoxicillin and ampicillin. Whilst increasing the diversity of the sample could be achieved by sampling greater numbers of PMEN1 isolates more widely, by studying a single lineage, increasing the sample size alone is unlikely to greatly reduce the effects of linkage disequilibrium- owing to the high level of relatedness between strains.

Finally, several gene candidates were identified, that carried SNPs statistically associated with changes in beta-lactam susceptibility. The *leuS* gene has not previously been identified as associating with clinical resistance, and therefore may offer a new insight into clinically circulating resistance. In addition, the cell wall synthesis pathway genes MurF and DdlA were found to similarly associate with changes in beta-lactam susceptibility. Whilst this is likely to reflect linkage,

due to selection occurring across the *pbp2b* gene (Enright and Spratt, 1999), it remains to be tested whether the activity of these genes is affected by this, and consequently whether they are involved in beta-lactam resistance

4 Isolating mitis group bacteria from the nasopharynx

4.1 Introduction

The rapid spread of multidrug resistance among pneumococci is typically attributed to the expansion of clonal lineages that have previously acquired genetic material encoding for drug resistance (Munoz et al., 1991, Munoz et al., 1992, Tomasz, 1997). *S. mitis*, and *S. oralis* are frequently cited as the sources of this resistance encoding material (Dowson et al., 1993, Sibold et al., 1994, Hakenbeck et al., 1998, Hart, 1998, Chi et al., 2007, Johnston et al., 2010). The nasopharynx comprises the pneumococci's primary niche (section 1.3.3). *S. mitis* and *S. oralis* have been found to be frequent colonisers of the nasopharynx among infants (≤2yrs) (Kononen et al., 2002). Furthermore, biofilm formation by the pneumococcus occurs within the nasopharynx, and consequently conditions facilitating a competent state should be met within the nasopharynx, rather than the linked oral cavity, from where pneumococci can occasionally also be isolated. Shared genetic material is found to occur between these species (Dowson et al., 1993). Importantly, sequences derived from *S. mitis* and *S. oralis* are found to occur within alleles encoding for reduced susceptibilities to beta-lactam antibiotics- drugs that are favoured for use in treating pneumococcal diseases (section 1.5). Whilst little is known of the recently discovered *S. pseudopneumoniae* (Arbique et al., 2004), the close relatedness of *S. mitis* and *S. oralis* to the pneumococcus is expected to facilitate the occurrence of recombinational exchanges between these three species (Lacks, 1966, Humbert et al., 1995).

Unlike the pneumococcus, which typically inhabits the host asymptomatically for periods of <6months before immune clearance (Hill et al., 2008), *S. mitis* and *S. oralis* are long-term colonisers of the oral cavities (Kononen et al., 2002). Their carriage lifestyle is thought to have exposed them to sufficient antibiotic pressures for *de novo* resistance to develop in these species. Clinical pneumococcal isolates that display reduced beta-lactam susceptibilities frequently possess sections of divergent sequence (i.e. section of sequence containing high SNP densities) within their *pbp* alleles. These sections are

furthermore found to contain SNPs that have been associated with a loss of beta-lactam susceptibility *in vitro*. Based on sequence homologies, these sequence blocks appear to have been derived from *S. mitis* and *S. oralis* isolates during inter-species recombination events (Dowson et al., 1993). Competence and fratricidal proteins possessed by the pneumococcus are thought to facilitate access to these resistance alleles (Whatmore et al., 1999). It is hypothesised that by acquiring resistance alleles via this mechanism, pneumococcal lineages have since exchanged this material with other pneumococcal lineages, facilitating global spread of resistance (Coffey et al., 1991).

These findings suggest that *de novo* beta-lactam resistance has not been the main driving force behind the spread of beta-lactam resistance among pneumococci. Resistance instead appears to have been acquired from other oral streptococci.

4.1.2 Classification of the oral streptococci

Colonisation of the oral and nasal cavities occurs soon after birth. A diverse microbiological flora exists within these sites and includes a number of streptococcal species (Aniansson et al., 1992). The oral streptococci have long been of interest for reasons such as their association with biofilm formation and maintenance (Tong et al., 2007), as indicators of a healthy microbiota (Preza et al., 2008), due to particular disease associations (Goldenberger et al., 1997, Claridge et al., 2001, Kadioglu et al., 2008), and have also yielded novel antimicrobials (Wescombe et al., 2011).

Classification of the oral streptococci has however proven difficult. During the history of this group several synonyms having been applied to describe the same organism (Kilian et al., 1989), and the group continues to be divided differently according to the character under scrutiny.

At their highest level the streptococcal genus can be divided according to their haemolytic properties, evident when grown on blood agar plates (Table 16).

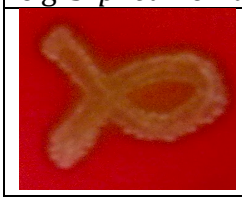
Alpha	**Beta**	**Gamma**
Partial haemolysis	Complete haemolysis	No haemolysis
e.g. *S. pneumoniae*	e.g. *S. pyogenes*	e.g. *Enterococci*
		

Table 16: Classification of the streptococcal genus was initially based on phenotypic properties observed when grown on blood agar plates. Images indicate the different levels of haemolysis observed when members of each group are grown on blood agar (R, 2014).

Lancefield (1933) was one of the earliest researchers to successfully develop criteria to distinguish between members of the haemolytic streptococci below this initial division. Lancefield's classification system was based on antibody-antigen agglutination reactions, and remains in use (Lancefield, 1933, Courtney and Li, 2013, Borst et al., 2013). The Lancefield system however was restricted to the beta haemolytic streptococci (with the exception of group F), which importantly excluded the pathogen *S. pneumoniae*. A lack of distinguishing characters among the remaining oral streptococci however complicated further classifications (Kilian et al., 1989). Genetic studies have since largely confirmed the inter-relatedness of this group, and have aided the development of the current classification scheme for these species (Aniansson et al., 1992).

The pneumococcus belongs to the mitis group, the largest of these assemblages, containing 8 species (Sakai et al., 2013)(Table 17). Although other members of this group occasionally cause disease (Goldenberger et al., 1997, Claridge et al., 2001), *S. pneumoniae* is the only member to be considered pathogenic. Its pathogenic status makes the pneumococcus the most researched member of the mitis group (Sakai et al., 2013). Most members of this group favour colonisation of the oral cavities. However, the connectedness between this and the nasopharynx means that other mitis group members can be isolated alongside pneumococci present in the nasopharynx (Kononen et al., 2002). Nasopharyngeal colonisation however does not appear to be stable among the streptococci, with carriage frequencies changing with age (Kononen et al., 2002).

Group	Species	Niche
Mitis group	*S. australis*	Mouth
	S. cristatus formerly *S. crista*	Mouth
	S. gordonii	Teeth, mouth
	S. infantis	Mouth and pharynx
	S. mitis	Mouth
	S. oligofermentans	Mouth
	S. oralis	Mouth
	S. parasanguinis (formerly *S. parasanguis*)	Throat
	S. peroris	Teeth and pharynx
	S. pneumoniae	Mouth and pharynx
	S. pseudopneumoniae	Mouth and pharynx
	S. sanguinis (formerly *S. sanguis*)	Mouth, dental plaque
	S. sinensis	Mouth
Anginosus group	*S. anginosus*	Mouth, throat, gastrointestinal and female urogenital tracts
	S. constellatus (subsp *constellatus*, *pharynges*)	Mouth, throat, gastrointestinal and female urogenital tracts
	S. intermedius	Mouth, throat, gastrointestinal and female urogenital tracts
Salivarius group	*S. salivarius*	Mouth, upper respiratory tract
	S. vestibularis	Mouth, upper respiratory tract
	S. thermophiles	Mouth, upper respiratory tract

Table 17: Members of three important oral streptococcal groups. The mitis group contains the important pathogen *S. pneumoniae*. Additional anginosus, and salivarius groups inhabit the oral cavities (Willcox et al., 2001, Sakai et al., 2013, Whiley et al., 1990).

The clinical significance of the oral streptococci means that accurate identification is of importance for both diagnostic and research purposes. Phenotypic and sequence based methods are widely available for identifying oral streptococci. However, their close relatedness, and similar morphological characters mean that these bacteria continue to be frequently misidentified (Kawamura et al., 1995).

Pneumococci are typically found to be optochin susceptible (14mm diameter zone of inhibition) (Arbique et al., 2004) and the pneumococcus capsule can be lysed following incubation in bile (2% sodium deoxycholate)(section 1.3.1). As such, optochin susceptibility and the "bile solubility test" are commonly used for the presumptive identification of pneumococci (Public Health England, 2014).

However, the size of the optochin inhibition zone is found to be variable, and different interpretations of how this test should be read have subsequently been suggested (Arbique et al., 2004). Furthermore, optochin resistant pneumococci (Borek et al., 1997, Pikis et al., 2001, Nunes et al., 2008), optochin susceptible viridans streptococci (Martin-Galiano et al., 2003) and bile insoluble pneumococcal forms have been identified widely (Obregon et al., 2002, Whatmore et al., 2000, Arbique et al., 2004).

To improve identification, principally of pneumococci, commercial kits based on biochemical profiling and serotyping have been developed. Biochemical kits have been favoured over serotyping based methods, as similar serotypes between species can lead to cross-reactivity among viridans streptococci (Arbique et al., 2004). Furthermore such tests fail to detect non-typeable pneumococci (isolates which cannot be assigned to a serotype), which have been estimated to account for up to 20% of the pneumococcal population (Cima-Cabal et al., 1999). Although biochemical profiling kits significantly reduce the timeframe (~4 versus ~72 hours) and labour requirements of conventional biochemical testing, they still lack specificity (Poutrel and Ryniewicz, 1984, Kawamura et al., 1999, Kikuchi et al., 1995, Arbique et al., 2004). Consequently these tests are often used in conjunction with optochin and bile susceptibility tests for the identification of oral streptococci (Arbique et al., 2004)(section 2.2, Figure 1).

Sequence based identification methods have largely failed to be successful in this area due to the existence of shared sequences across this group. Although the 16S gene is often targeted in bacterial profiling studies, Kawamura and colleagues (1995) demonstrated sequence homologies of greater than 99% existed between *S. mitis, S. oralis* and *S. pneumoniae* species.

A more recent focus has been on the virulence determinants possessed by *S. pneumoniae,* which are thought to be essential for its pathogenic lifestyle during colonisation and survival within the host (Hirst et al., 2004, Kadioglu et al., 2008).

Gene probes that targeted the autolysin encoding gene *lytA*, and associated pneumolysin (*ply*) toxin were initially used successfully for this purpose (Diaz et al., 1992, Hassan-King et al., 1994, Cima-Cabal et al., 1999, Dagan et al., 1998, Whatmore et al., 1999). However the usage of these targets for taxonomic purposes has since subsided following the identification of these genes among other members of the oral streptococci (Whatmore et al., 2000, Poyart et al., 1998, Kawamura et al., 1999). These findings suggest that although virulence genes may not be required by other streptococcal species, these genes are still on occasion exchanged between the oral streptococci. Gene probes based on other genes, such as resistant *pbp* alleles have also suffered from a lack of specificity owing to shared genetic material across this group (Dowson et al., 1993, Arbique et al., 2004).

Consequently researchers have attempted to identify conserved sets of species-specific genes (Maiden et al., 1998, Arbique et al., 2004). Although this approach has so far proven to be largely successful, this method is not suitable for use in most diagnostic settings. This is largely due to cost, and labour constraints (Arbique et al., 2004). Consequently, gene probe methods remain commonly in use despite their pitfalls.

It is currently unknown whether antibiotic pressures have been sufficient for resistance to develop within viridans group streptococci in Malawi and to what extent these bacteria have contributed to beta-lactam resistance among Malawian pneumococci. Furthermore little is known about how frequently these species come into contact, and consequently what opportunities exist for genetic exchanges to take place between these species. Changes in drug usage following the introduction of ceftriaxone into Malawi are also likely to have lead to a change in selection pressure since it's introduction in 2003, and increased usage in 2007. However, owing to the difficulties associated with identifying different members of the streptococcal group, very few attempts have been made to characterise the relationship between the viridans streptococci and pneumococci present in the nasopharynx.

The aim of this study was to investigate the co-occurrence *S. mitis*, *S. oralis* and *S. pneumoniae* within the nasopharynx over a period of different antibiotic regimes.

4.1.3 Study Aims

- To develop a high-throughput method for isolating *S. oralis* and *S. mitis* from nasopharyngeal samples
- To determine the frequency with which mitis group species come into contact in the nasopharynx

4.2 Study samples

Staff at QECH routinely take CSF and blood samples for culture. However nasopharyngeal samples necessary to investigate bacterial carriage are only available from studies where specific ethical approval was sought. This sampling method was approved for two studies, one undertaken from 2003 until 2009 (VacServ, ethical code: NHRSC 1073), and the second from 2010 onwards (Flu surveillance, ethical code: P.07/10/958). In total both studies collected approximately 600 nasopharyngeal samples between 2003 and 2012, the study period considered.

Each of these studies used a different sampling protocol. Per-nasal swabbing (section 2.2.3) was undertaken in the VacServ study. Here, a per-nasal swab is passed through the nostril until it reaches the anterior nose, before being removed and stored in a vial of STGG broth (O'Brien et al., 2003, Gibson and Khoury, 1986).

In contrast, the Flu surveillance study collected nasal aspirates using a suction catheter in accordance with section 2.2.4. During this procedure 1 to 2ml of sterile saline was instilled into the patients nostril and a catheter tube threaded through the nostril until it reached the nasopharynx. The aspirate was then collected using the suction catheter, with the procedure being repeated until sufficient volume of sample was collected. In the Flu surveillance study, samples

were stored in duplicate UTM and STGG vials. The Flu surveillance study began in late 2010, with collection on going.

In addition to the clinical samples described above, control strains were obtained from the MLW diagnostics team (Table 18).

Species	Control used
S. pneumoniae	NCTC 12977 (ATCC 49619)
S. constellatus	MM9889a, NCTC 13122
S. mitis	NEQAS*
S. oralis	NEQAS*

Table 18: Control strains used. *strains provided by the National External Quality Assessment Services (NEQAS).

4.2.1 Isolating mitis bacteria *in vitro*

Pneumococci were identified in accordance with the protocols used by the MLW diagnostics staff. Briefly, samples vials were chosen at random from each year of the study period. Samples were thawed and vortexed, before a sterile loop was used to inoculate 10µL of sample onto a BHI agar plate supplemented with 5% sheep blood and 5µg/mL of gentamycin. Gentamycin inhibits the growth of nasopharyngeal co-colonisers, allowing pneumococci to be more readily detected. A 5µg optochin disc was also placed on the plate, optochin susceptibility being indicative of *S. pneumoniae* (CDC, 2012). Colonies identified as being alpha haemolytic, and optochin susceptible were presumptively identified as *S. pneumoniae* using this method (section 2.3.3).

Pneumococci were isolated from 65% of the samples tested, with the study type appearing to affect the ability to recover pneumococci. Under the VacServ per-nasal swabbing method pneumococci were isolated from 75% of samples tested, whereas pneumococci were only isolated from 52% of samples collected under the Flu surveillance study. This suggests that the per-nasal swabbing method was either more effective at sampling the nasopharynx, or that pneumococci were better preserved using this method.

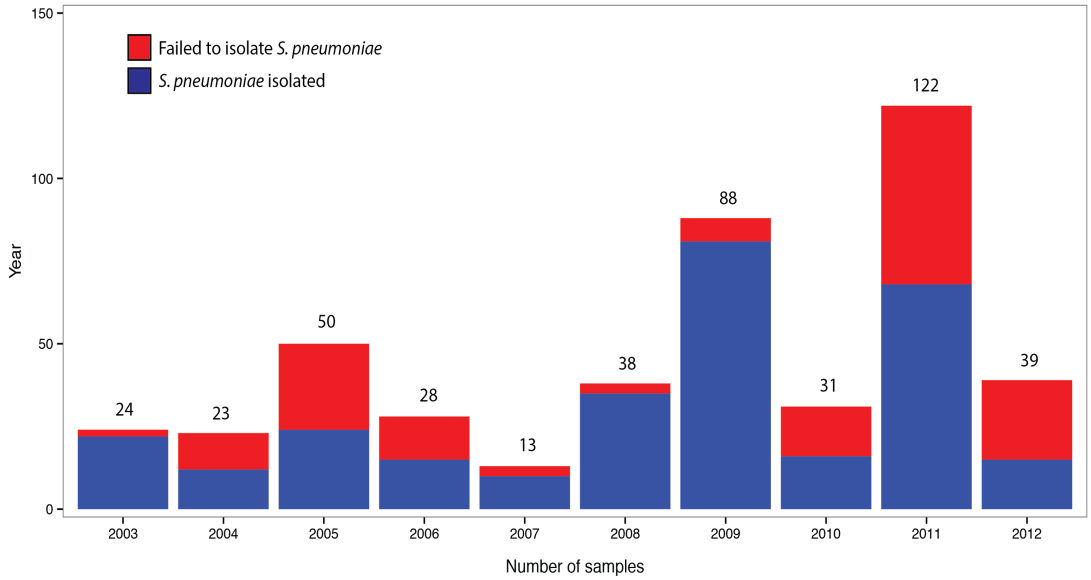


Figure 41: Recovery of pneumococci from nasopharyngeal samples per year, coloured according to whether *S. pneumoniae* could be isolated from the sample (blue), or not (red). Numbers above each bar in the graph indicate the total numbers of isolates tested for each year. Pneumococcal recovery can generally be seen to be greater proportionally within the VacServ samples (2003-2009), compared to the Flu surveillance samples (2010-2012). Fewer samples were available for testing in the VacServ study, which can also be seen from the graph.

Whilst pneumococci are routinely isolated for diagnostic purposes, other members of the mitis group are not. Presumptive identification of viridans group species typically relies on identifying optochin resistant colonies, which have been otherwise determined not to be pneumococci, when grown on blood agar plates (Arbique et al., 2004). In a diagnostic setting a commercial biochemical profiling kit is often used in addition to confirm species identity. Such kits were unsuitable for a high throughput approach due to cost.

Blood plates, lacking gentamicin however, were found to be unsuitable for the isolation of these species from nasopharyngeal samples. Nasopharyngeal samples grown on these plates became overgrown with non-streptococcal species. Staphylococci appeared to commonly occur in these samples and overgrow the plate surface. To suppress the growth if other colonisers, an alternative growth media, Mitis Salivarius Agar (Acumedia), supplemented with 1% tellurite was opted for. This media was specifically developed to differentiate between species of *Streptococci* and *Enterococci* while inhibiting the growth of *Escherichia* and *Staphylococcus* species (Chapman, 1946, Facklam, 1991, Takada et al., 2006). In addition to suppressing the growth of these co-colonising species, Mitis Salivarius Agar also gives colonies of interest particular morphological

characteristics, aiding identification. This is in part achieved through the action of trypan blue, which is present in the media and is absorbed into colonies growing on the plate. For the isolation of mitis group species, plates were incubated in accordance with the manufacturers instructions, for 24 hours before being inspected for colonies displaying morphological characters consistent with being part of this group. When grown on this media enterococci are characterised as producing blue-black colonies, whilst viridans group streptococci typically result in lighter, blue colonies (Figure 42)(delrio.dccd.edu).

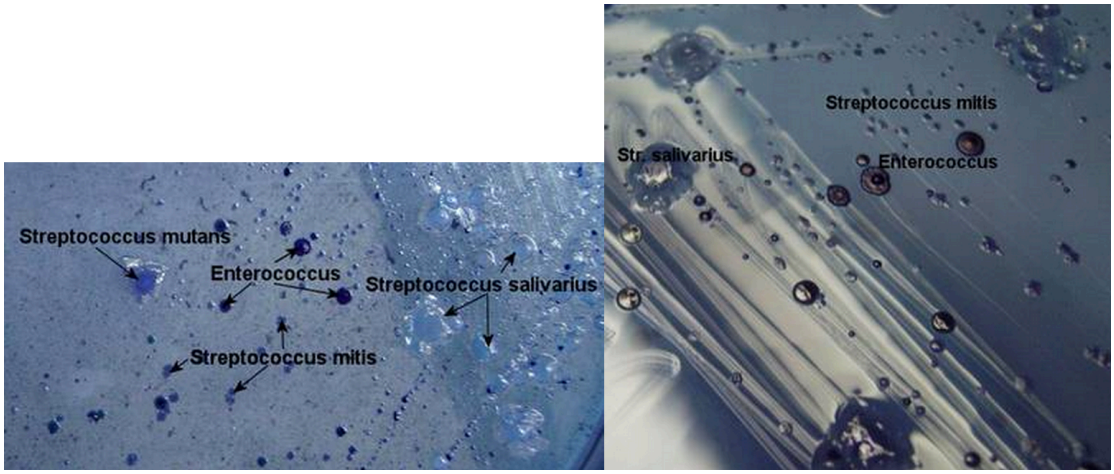


Figure 42: Colony morphologies when grown on Mitis Salivarius Agar for Enteroccoi and select Streptococcal species (delrio.dccd.edu).

Following this approach, colonies that fulfilled the morphological descriptions for mitis group species (small, light blue, flat colonies; delrio.dccd.edu) were regrown on blood agar plates to determine optochin susceptibility, catalase activity and gram staining. This additional testing was to ensure *S. mitis* and *S. oralis* species had been isolated; mitis group members being catalase-negative, and optochin resistant (CDC, 2012).

In total 456 samples were grown on Mitis Salivarius Agar. Isolates conforming to the criteria for mitis group streptococci were subsequently tested, but no members of this group were identified based on the above protocol.

The inability to isolate *S. mitis* and *S. oralis* from the nasopharyngeal samples suggested that either these species could not be sampled from the nasopharynx

alongside pneumococci, or that the subsequent storage and isolation regime was not suitable for the recovery of these species.

In both the VacServ and Flu surveillance studies STGG media (appendix 10.5) was used to store the samples. Whilst this media is recommended for the storage of pneumococci from nasopharyngeal samples, it is untested whether this media is similarly favourable for the storage of other mitis group species (O'Brien et al., 2001). Furthermore, STGG samples were subject to overnight incubation at 37^{o}C 5% CO_2 in order to increase subsequent pneumococcal recovery (pers. comm. Everett 2012). Consequently, growth competition could also have affected the recovery of these species.

To determine whether STGG media supported the isolation of other oral streptococci, control strains representing oral streptococci were cultured overnight in STGG broth. Controls included *S. oralis,* and *S. mitis*, as these were principally of interest to this study. In addition, *S. constellatus* was included, as this was representative of the Anginosus group, which similarly commonly inhabits the oral cavities (pers. comm. Wilson 2012). To investigate whether competition between pneumococci and these other species affected their recovery, each species was co-cultured with pneumococci in STGG broth. Inoculums were prepared by resuspending culture from overnight plate growths in PBS to a turbidity equivalent to a 0.5 McFarland standard.

To test whether STGG media supported growth of these species, each vial containing STGG media was inoculated with 10µL of the suspension, and incubated overnight as described above. To determine whether pneumococci could outcompete the growth of these other species in STGG media, vials were inoculated with pneumococci and the additional species being tested. Equal volumes of each suspension were used in each case, although the total volume added to the STGG broth did not exceed 10µL.

Following overnight growth, STGG media was inoculated onto a blood agar plate and grown overnight. Pneumococcal presence was subsequently determined

based on colony morphology (being larger than the other species tested), and optochin susceptibility. PCR was used to determine if the other streptococcal species were present or not (see Table 19 for primers).

Species	**Target**	**Sequence**	**Amplicon length**
Anginosus group	*pbp2b*	5′-TGCTGCAACGGTAGCTAATGG-3′ 5′-CAAAGGTTTCTGCTGTCCCTG-3′ (Takao et al., 2004)	275bp
S. mitis	*pheA*	5′-TGGCTTATC CTTCCTAGATGG-3′ 5′-GATTGCGGTCGACAAA-3′ (Park et al., 2012)	557bp
S. oralis	*rgg*	5′-GC TTTGACCGAACAGTTTCC-3′ 5′-CATTGG TATTCCCCACCTTG-3′ (Park et al., 2010)	475bp

Table 19: Primer sequences used to identify *S. mitis*, *S. oralis* and anginosus group bacteria. Reference strains were chosen from the MLW diagnostics archive, and *S. constellatus* was used as a representative strain for the anginosus group. See section 2.4 for PCR conditions and protocol.

The results from this study (Table 20) indicated that STGG media did support the growth of other oral streptococci. Furthermore, in no case was one species out-competed to the extent that it could not be detected. This study was not quantitative however, and so it is not possible to determine to what CFU each species grew to following incubation.

Assay	**Result**
S. constellatus	Present
S. constellatus x *S. oralis*	All present
S. constellatus x *S. pneumoniae* x *S. oralis*	All present
S. pneumoniae x *S. oralis*	All present
S. oralis	Present
S. pneumoniae x *S. constellatus*	All present
S. mitis	Present
S. mitis x *S. oralis*	All present
S. mitis x *S. pneumoniae* x *S. oralis*	All present

Table 20: Present/absence results from the growth of single and co-cultured growth experiments.

Whilst no competitive effect was identified using the above approach, the starting volumes used were not chosen to represent inoculums typically found for nasopharyngeal samples, which is presumably variable depending on the

patient. Consequently, a competitive affect could still be present, either owing to the presence of other nasopharyngeal colonisers not tested here, or due to differences in prevalence of species in the nasopharynx. An alternative PCR based approach was therefore carried out.

For samples collected under the Flu surveillance study, once collected the aspirate was divided into two UTM samples in addition to two STGG vials. The UTM samples were not subject to overnight growth, but were stored immediately at -80^{o}C. As such, although the UTM media did not support bacterial survival, the genetic material present following original aspirate collection was stored in these samples. If viridans group streptococci were present in these samples, but outcompeted during subsequent overnight incubation, they would still be detectable by PCR in these original samples. Hence this approach should be more sensitive to species detection. PCRs carried out on the raw UTM samples however did not identify *S. oralis, S. mitis* or *S. constellatus* as being present. This further supported the finding that although the pneumococcus is thought to frequently colonise the nasopharynx alongside other oral streptococci, these species are not readily isolated alongside each other. This suggests the opportunity for genetic exchange between these species does not occur frequently.

To determine the ability to detect mitis group bacteria from mixed samples from other oral cavities, 13 volunteers were asked to undergo a mouth swabbing. In each case the mouth swab was first streaked onto a blood agar plate, and the swab subsequently placed in PBS to resuspend the remaining sample. The plates were then incubated overnight, whilst a PCR was carried out on the PBS suspension to determine whether there was a difference in the ability to detect these species based on whether the sample was incubated overnight or not. The PCR primers used included those for anginosus group species, *S. mitis* and *S. oralis*.

Species	Sample form	Recovered
Anginosus group	PBS	0
	Plate	4
S. mitis	PBS	5
	Plate	1
S. oralis	PBS	6
	Plate	2

Table 21: PCR results for each of the different target species based on oral samples. Species/group is indicated in the left column. Sample form indicates whether the PCR was conducted on the raw swab sample, or following overnight plate culture. "Recovered" column indicates the number of samples from which each particular species/group was identified.

Table 21 indicates counts of the number of volunteers testing positive for each species of interest under different sample preparations. The PBS suspension represents samples tested immediately following sampling. In contrast, plate samples were incubated overnight following streaking of the raw sample onto a blood agar plate. Consequently, these samples would have been subjected to resource competition between strains. This experiment was intended to determine whether there were differences between the species detected depending on how the sample was prepared. Whilst there is some indication that *S. mitis* and *S. oralis* were more successfully isolated from the raw sample than those grown overnight on blood agar, it is also clear that viridans streptococci were readily isolated from oral samples. This suggests that although *S. oralis* and *S. mitis* are frequent colonisers of the oral cavity, they rarely occur in the nasopharynx. This also indicates that the methods used in this study previously were sufficient to identify these species. This supports the initial conclusion that these species rarely come into contact with pneumococci.

As such, a laboratory approach to identifying such co-occurrences, without prior knowledge that a sample contains either *S. mitis* or *S. oralis* may be unsuccessful for two reasons. Firstly, large numbers of samples will be grown unnecessarily as they do not contain *S. mitis* or *S. oralis,* and secondly, if these two species are frequently outcompeted during normal culture, then the laboratory growth and isolation protocol may need to be adjusted to reflect this- for example with repeated growth of the same sample to maximise the chance of identifying and isolating either *S. mitis* or *S. oralis*.

An alternative approach would be to carry out PCR screens routinely on the raw samples immediately following collection. However, this is resource and time consuming, and consequently would be unsuitable for most diagnostics settings. Another approach which could reduce the time and cost of PCR screening would be to identify conserved *S. oralis* and *S. mitis* within a single shared locus, such such as the 16S gene. The 16S gene has been favoured for taxonomic classification, as it has been found to be conserved across all bacterial species studied, contains conserved regions, which can be targeted by universal primers, and contains variable regions, which can be used for species identification (Woese, 1987, Van de Peer et al., 1996, Vinje et al., 2014). The current method relies on using universal primers to amplify a section of the 16S gene, and has been found to be widely successful at distinguishing between bacterial species (Janda and Abbott, 2007). Despite this, Kawamura and colleagues (1995) have shown that sequence homologies exist at levels greater than 99% between *S. pneumoniae, S. mitis* and *S. oralis* when comparing similarities across the entire 16S gene. However, commercial gene probes (AccuProbe) have been developed that claim to successfully differentiate between these species based on specific regions of the 16S gene (Carvalho et al., 2003, Whatmore et al., 2000, Arbique et al., 2004). Consequently it was hypothesised that species-specific SNP patterns may exist within the 16S gene, which are masked when comparing 16S sequences over their entire lengths (Kawamura et al., 1995).

This study attempted to identify species-specific SNP patterns from publicly available 16S sequence reads *in silico.* The aim of which was to subsequently use these "markers" to identify species of the mitis group from 16S reads collated from clinical nasopharyngeal samples.

The clinically sampled 16S reads available for analysis had been sampled from QECH, Malawi. Although the raw samples were no longer available for *in vitro* characterisation, the intention was to provide a real-world dataset to test any species-specific markers identified, and in addition offered an alternative approach for identifying the occurrence of *S. mitis* and *S. oralis* from nasopharyngeal samples.

Given the close proximity and connectivity between the nasopharynx and the oral cavity, representative sequences from all oral streptococci were included (Table 22). This ensured that SNP patterns could be compared across the oral streptococcal group, allowing verification of whether these were specific to one species or group.

Sequence reads were downloaded from the publicly available 16S sequence archive http://www.arb-silva.de/. This database was chosen as reads are quality checked prior to inclusion. This includes taxonomic verification, and consequently should reduce the occurrence of erroneously assigned streptococcal group reads being included (Silva, 2015). In addition the 16S sequence for a non-typeable pneumococcus (ST 334 NT lineage) was included. This isolate was included as nontypeable pneumococci are frequently found in carriage, and possess atypical pneumococcal characters (Salter et al., 2012). Importantly, unencapsulated pneumococci are thought to recombine at a higher rate than encapsulated forms (Chewapreecha et al., 2014a), offering greater opportunity for genetic exchanges to affect the 16S gene. As such it was important that any species specific SNP patterns identified were also assessed relative to this isolate, which is currently the only confirmed nontypeable pneumococcal isolate for which 16S sequence is available (pers. comm. Salter 2013).

The initial dataset contained 493 16S reads across the three streptococcal groups considered. The majority of 16S reads belonged to members of the mitis group, which also contained the most taxa. In addition, the mitis group contains *S. pneumoniae*, for which 139 16S reads were available (Table 6). The length of the 16S amplicon was variable between species ranging from 1252bp to 1568bp. Therefore, to aid comparison, reads were first aligned using ClustalO. The quality of the alignment was then checked manually before a maximum likelihood phylogeny was constructed using RAxML.

Group	Species	Number of reads
Mitis group (Total= 314, 175, excluding *S. pneumoniae*)	*S. australis*	6
	S. cristatus formerly *S. crista*	8
	S. gordonii	15
	S. infantis	17
	S. mitis	34
	S. oligofermentans	2
	S. oralis	29
	S. parasanguinis (formerly *S. parasanguis*)	25
	S. peroris	3
	S. pneumoniae	139
	S. pseudopneumoniae	9
	S. sanguinis (formerly *S. sanguis*)	25
	S. sinensis	2
Anginosus group (Total= 77)	*S. anginosus*	37
	S. constellatus (subsp *constellatus*, *pharynges*)	27
	S. intermedius	13
Salivarius group (Total= 102)	*S. salivarius*	65
	S. vestibularis	2
	S. thermophilus	35

Table 22: Taxa included in the 16S analysis. Total counts for members of the group are indicated in the left hand column, while total numbers of reads present for each taxa are indicated in the right column.

The phylogenetic analysis of the 16S gene showed that species generally clustered well at the group level. All members of the salivarius group clustered together. However, members of the anginosus group showed less fidelity, with a small group bisecting the mitis cluster. A small number of mitis group members also associated away from the main mitis group cluster, forming a small group at the base of the tree (Figure 43).

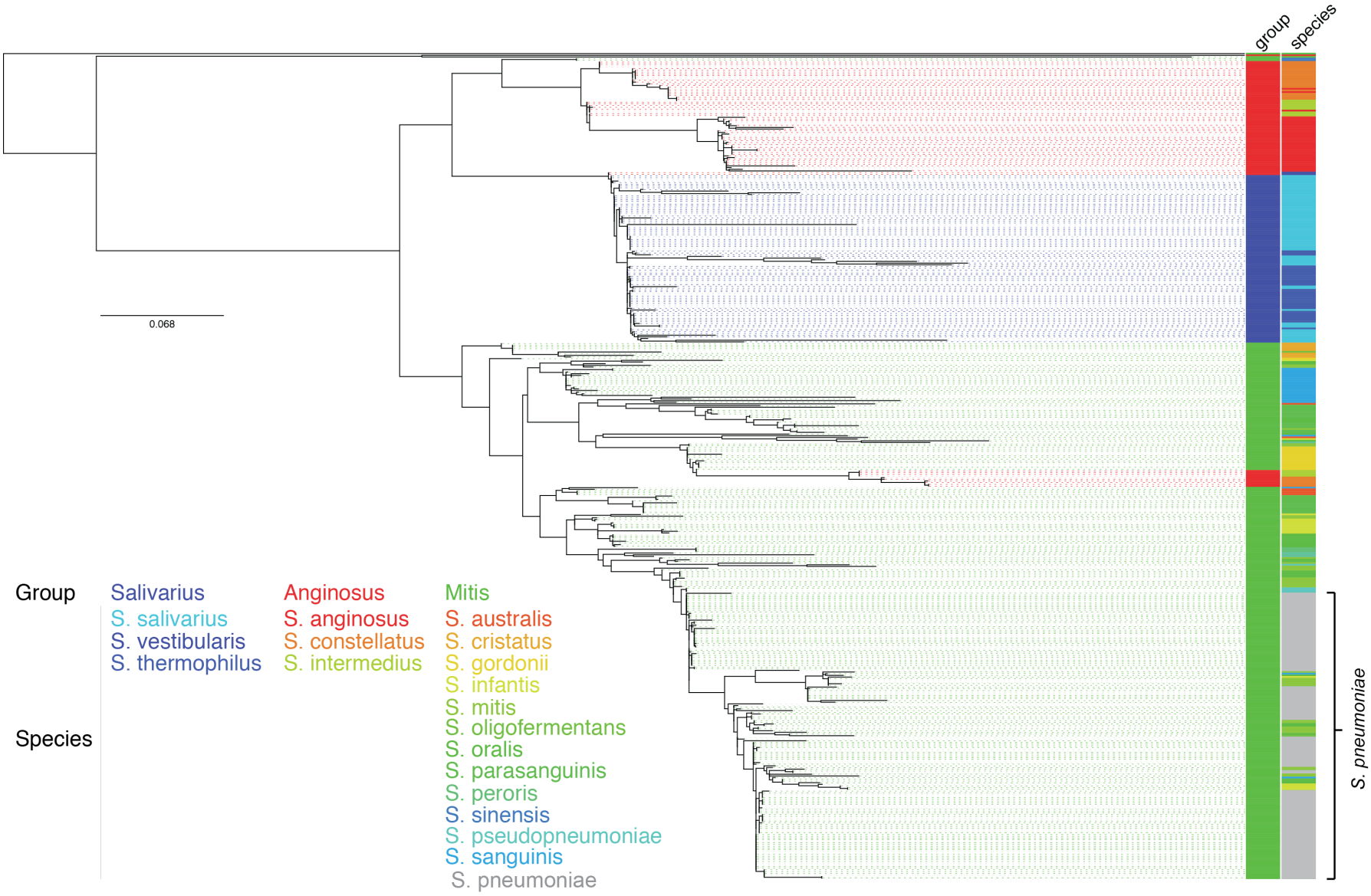


Figure 43: Phylogenetic tree constructed based on nucleotide variation present in the Streptococcal 16S reads. Group (anginosus, salivarius, mitis) is marked in the first column. The species is coloured in the second column.

There appeared to be little sequence conservation at the species level, across all groups except for *S. pneumoniae*. Although this group was interspersed with three distinct clusters that included other mitis group species, branch lengths were visibly shorter, indicating high levels of sequence conservation. This may in part be due to denser sampling of pneumococcal isolates, although others have suggested levels of diversity are more restricted among pneumococci than among other mitis group members in particular (Kilian et al., 2014).

To visualise the variation present with respect to pneumococci, SNPs were identified by comparison with ATCC700669. This isolate was chosen as a reference owing to it being positioned within the middle of the pneumococcal cluster, and due to the high level of annotation available for this isolate (Croucher et al., 2009). On comparison with this sequence, 1045 SNPs were identified across all 493 isolates compared. In some cases species specific SNPs were identified. For example a single SNP in *S. anginosus* was present among 32 of the 37 isolates tested. However, no SNP, or groups of SNPs were found across all members of a single species, as such it was not possible to reliably identify species based on conserved SNP patterns alone.

At the group level 8 SNPs were present in all but one salivarius isolate, but not found in any other oral streptococcal group compared. In addition 4 SNPs were present across all isolates of the anginosus group tested, but were absent from all other oral streptococcal groups tested. This suggested that these SNPs could be used to differentiate isolates at the group level. In addition, 722 SNPs were identified as occurring solely in the 354 non-pneumococcal species. This suggested that anginosus group and salivarius group isolates could be distinguished based on conserved SNPs. In addition, there was also the possibility that particular SNPs may occur uniquely in non-pneumococcal streptococcal species. These hypotheses were tested further to determine whether these SNP markers could be used to identify the presence of non-pneumococcal species from nasopharyngeal samples- represented by the Malawian 16S reads.

The Malawian 16S reads represented nasopharyngeal samples collected from 58 subjects, 40 children, and 18 adults. These reads were amplified from the V3-V5 region of the 16S gene, which contains nine variable regions in total (V1-V9) (Van de Peer et al., 1996). This region was amplified using PCR, before being sequenced on a Roche/454 platform. The *in silico* sequences were then curated manually, to remove incorrect assemblies, resulting from PCR and sequencing errors (Kamng'ona, unpublished 2015).

As the Malawian 16S reads were collected from several nasopharyngeal swabs, a diverse selection of species was likely to be present, broadly representative of the nasopharyngeal community present in Malawi. The search space was reduced compared to the sequences available from www.arb-silva.de, as the universal 16S primers used on the nasopharyngeal samples targeted an approximately 400bp internal segment of the entire gene, the V3-V5 region- although the SNP markers identified previously did occur in the amplified region. The V3-V5 region is commonly chosen for amplification in order to increase the overall read length across the variable 16S regions. This allows a higher-throughput than would be achieved by sequencing the entire gene, but is similarly of sufficient length to aid subsequent taxonomic differentiation.

In total 173 reads were available in the Malawian 16S dataset. These reads were first aligned to the reference ATCC700669 16S gene using ClustalO, and the SNPs present were identified relative to this reference sequence. This ensured that the SNP positions identified were directly comparable to the positions of the SNP markers being compared. This alignment was then used to construct a maximum likelihood phylogeny, so that read divergence could be visualised.

From the phylogenetic analysis, 3 clear clusters (coloured and labelled; Figure 44) emerged with the pneumococcal ATT700669 reference occurring in cluster 2, and a further mixed group was present towards the terminus of the tree. This clustering pattern implied sequence reads within these clusters may be closely related at a taxonomic level. In order to test this observation, several (≥5) representative sequences were chosen across each cluster and submitted to

BLAST. This allowed similar archived sequences to be compared. Sequence information available for the most similar species was then used to infer the likely taxonomic identity of the 16S read being compared. Using this approach, clusters were found to contain a broad range of bacteria, which in the majority of cases belonged to the same phyla, indicating some degree of relatedness occurred within these clusters. However, such levels of similarity only occurred at the level of phyla, confirming that the Malawian 16S reads were highly diverse in terms of the inferred species present.

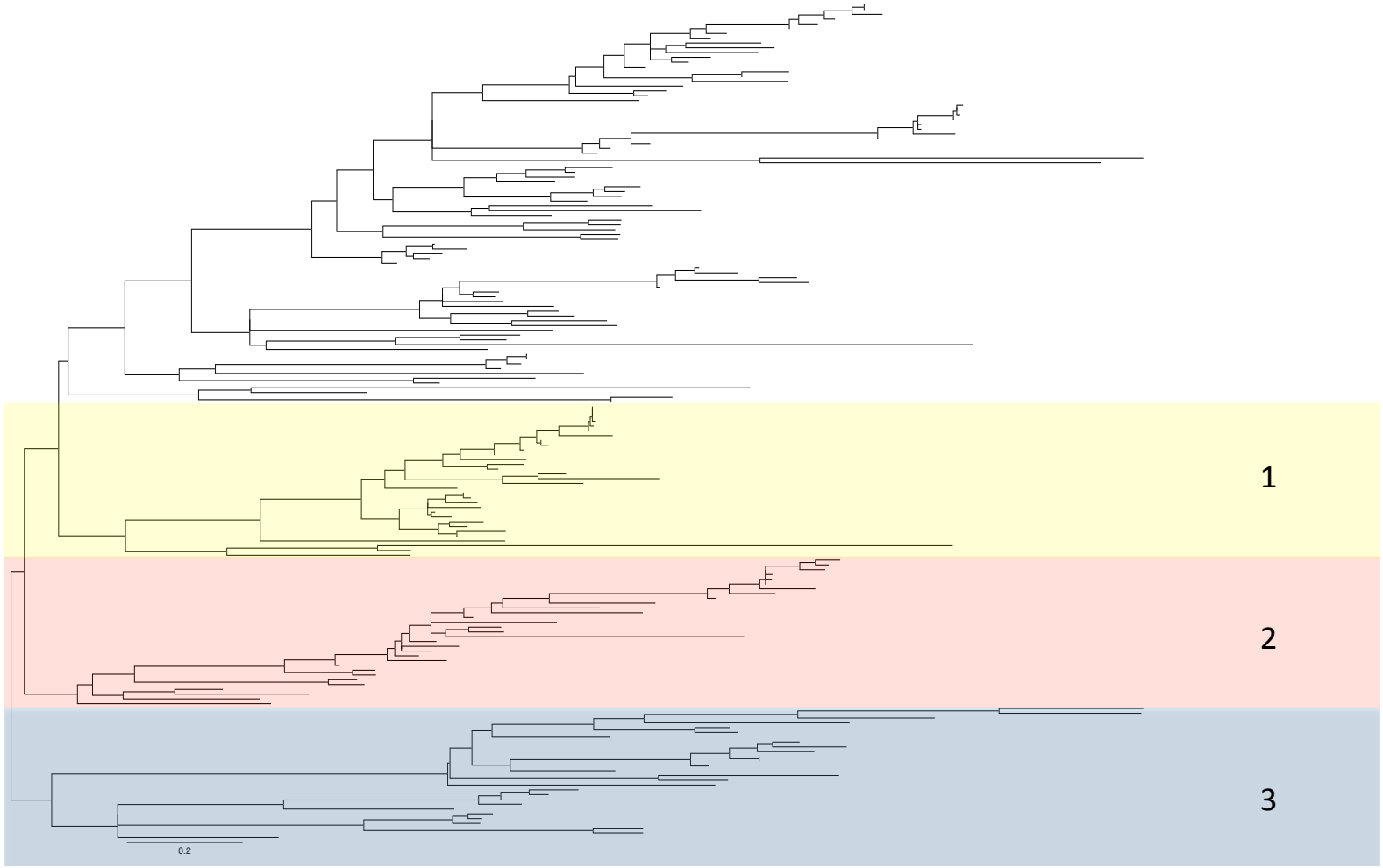


Figure 44: Phylogenetic tree constructed based on publicly available 16S sequences for viridans group streptococci, and *S. pneumoniae*. Three obvious clusters (coloured and numbered) appeared, with a mixed group towards the terminus of the tree. The reference sequence occurred at the terminus of cluster 2.

Two comparisons were performed in order to determine the ability of the SNP markers to differentiate the species present in the Malawian 16S reads. Firstly, the SNPs identified across all non-pneumococcal streptococci were compared to the SNPs present in the Malawian dataset. This comparison revealed that all 252 SNP markers occurred in the MLW sequence reads, but that these were present throughout the tree. Furthermore, using BLAST, Malawian reads that shared these SNPs belonged to a diverse range of taxa. Consequently, these SNP markers were not successful at distinguishing species at this level of diversity. In the second comparison, the SNP markers unique to the anginosus and salivarius groups were both compared to the SNPs present among the Malawian 16S reads. This comparison similarly indicated that these SNPs were present in non-streptococcal taxa as identified through sequence comparison in BLAST. Consequently, the SNP markers did not allow confident identification of streptococcal species from mixed 16S reads at this level of sequence diversity.

Using BLAST, the reads present in cluster 2, which contained the pneumococcal reference sequence were presumptively assigned to species. This indicated that only the 6 sequences immediately flanking the reference ATCC700669 sequence belonged to the streptococcal genus. In addition, SNPs markers present among non-pneumococci had been identified in these isolates (Table 23). To better view the variation present in these putatively assigned streptococci, a phylogeny was constructed based on these reads alone (Figure 45). SNPs present were plotted alongside, relative to their positions in the 16S gene and coloured according to whether they had been identified as occurring among non-pneumococcal species (red) or not (grey). In total 44 SNPs were identified based on comparison with the reference, with 13 of these previously identified as occurring in non-pneumococcal groups isolates Three isolates were identified as containing clusters of SNPs at a position approximately 800bp into the 16S gene. However, few of these SNPs were present in the non-pneumococcal marker set, and the BLAST results indicated these reads most likely belonged to pneumococci depite the presence of the SNP clusters.

ID	pubmed ID	Query coverage	Identity	Non-pneumococcal SNPs	Total SNPs in sequence	Putative species
Read1	HBJUJ2Q02JDSJW	0.89	0.99	1	2	*S. pneumoniae*
Read2	HBJUJ2Q02H4RWD	0.86	0.99	0	2	*S. pneumoniae*
Read3	HBJUJ2Q02G7FHX	0.86	0.99	1	15	*S. pneumoniae*
Read4	HBJUJ2Q02ILR9D	0.83	0.96	1	13	*S. pneumoniae*
Read5	HBJUJ2Q01A442C	0.89	0.96	2	16	*S. pneumoniae*
Read6	HBJUJ2Q01BVMRB	0.89	0.99	10	20	*S. dysgalactiae*

Table 23: Species identification based on BLAST. Only six reads could be identified as belonging to the Streptococcal family using this approach- with all but one read appearing to be derived from *S. pneumoniae*.

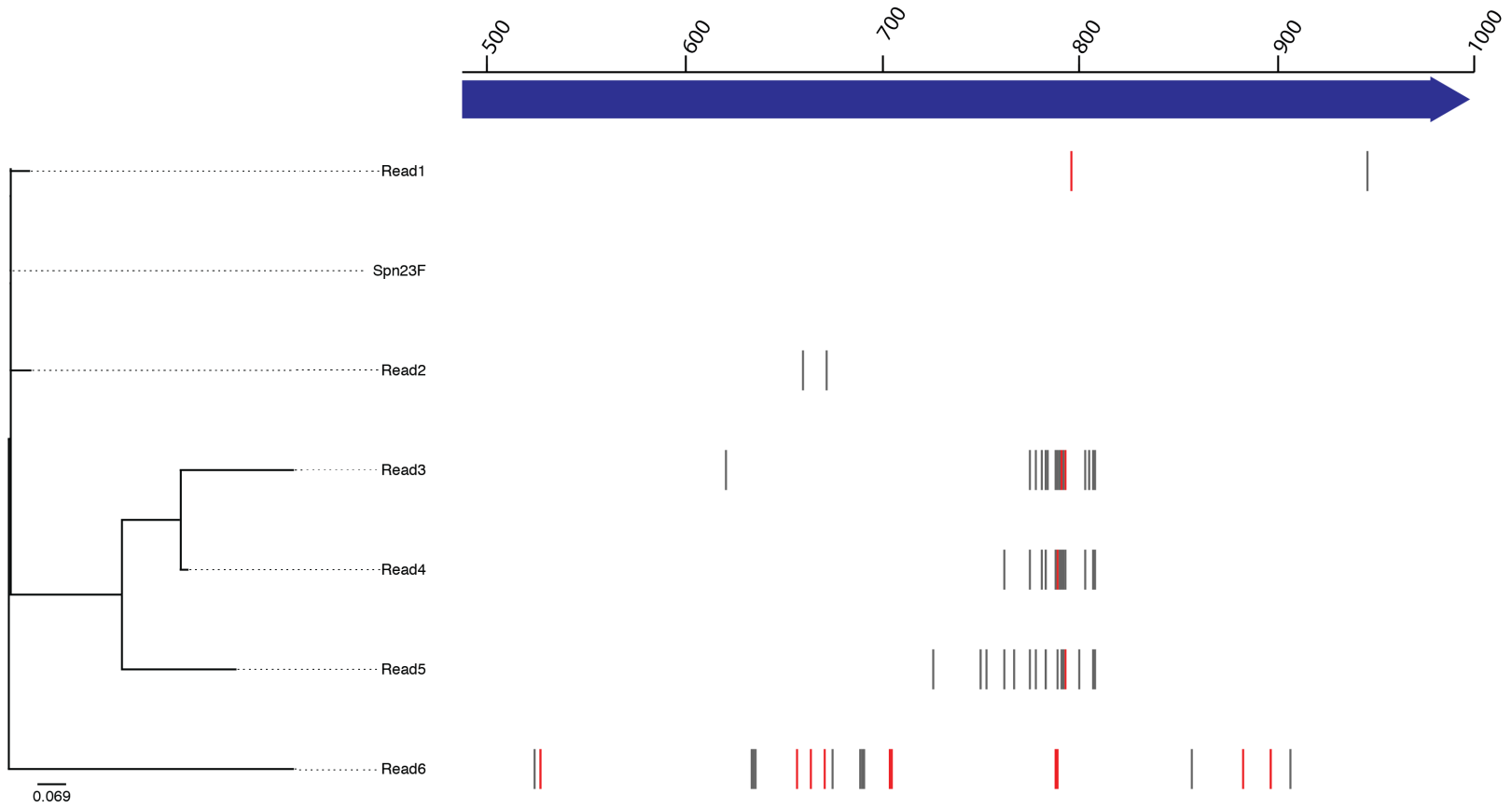


Figure 45: Phylogenetic tree constructed based on 16S variation present for the six reads identified as belonging to Streptococcal species. SNPs are plotted alongside, relative to their position in the gene. SNPs coloured red were identified previously as unique to non-pneumococcal isolates.

Read6 was a clear outlier based on the phylogeny, and contained a relatively large number of SNPs distributed throughout this region of the 16S gene. In addition 10, out the 18 SNPs identified in this isolate had previously been identified as occurring among non-pneumococcal streptococci. This was consistent with the BLAST result, which identified this read as most similar to *S. dysgalactiae* susp. *equisimilis. S. dysgalactiae*, is a beta haemolytic streptococcus which is commonly associated with Lancefield groups C and G (Jensen and Kilian, 2012). This species is however more divergent than the other streptococcal species included in this analysis, indicating that genetic exchange between pneumococci and this species would likely be limited due to a lack of sequence homology (Humbert et al., 1995).

This analysis appears to highlight the difficulty of differentiating streptococcal species using 16S data. No SNP markers could be identified in this study that could successfully delimit between members of the streptococcal genus to the species level. This could be an artefact of the 16S dataset used here only representative of the V3-V5 region of this gene however. Furthermore, whilst sequence similarity using BLAST was used to infer species identity for 16S reads, as the nasopharyngeal samples for the Malawian 16S reads were no longer available, it was not possible to confirm these results *in vitro.*

In summary, the V3-V5 region of the 16S gene appeared a poor target for the differentiation of mitis species. It remains to be tested whether other regions of the 16S gene could be used for this purpose however. Consequently, PCR screening using the primers developed by Park and colleagues (2008, 2010) would offer a more preferable method for identifying *S. mitis* and *S. oralis* from samples in which these species co-occur with *S. pneumoniae.*

4.3 Discussion

The nasopharynx is widely reported as the site of genetic exchange between *S. mitis, S. oralis* and *S. pneumoniae*. However, in this study both *in vitro* and *in silico* attempts to identify *S. mitis* and *S. oralis* from nasopharyngeal samples were

unsuccessful. This suggests that the co-occurrence of these species with pneumococci in the nasopharynx is a rare event. Assuming that the nasopharynx is the primary site for recombination between these species, the results presented here indicate that the co-occurrence of these species with pneumococci occurs infrequently, and that pneumococci therefore have limited opportunities to acquire resistance-encoding genetic material from these species. Consequently, although beta-lactam resistance may be widely present among the viridans streptococci, this source of resistance encoding material may not be as accessible to pneumococci as has been previously assumed. Furthermore, it is possible that only under certain conditions do pneumococci and these species co-occur.

An individual's age likely plays an important role in this process. Kononen and colleagues (2002) studied the establishment of streptococci in the oral and nasopharyngeal cavities of infants (≤2yrs). This study, in accordance with the results presented here, showed that *S. mitis* and *S. oralis* were frequently identified from oral samples (saliva). In addition these authors did identify *S. mitis* and *S. oralis* in nasopharyngeal samples, citing them, along with pneumococci, as the most frequent colonisers of this site. The ability of all three species to circumvent the hosts immune system through cleavage of IgAI antibody appeared important for initial colonisation of both oral and nasopharyngeal sites (Kononen et al., 2002). However, immune clearance appears to be an important factor preventing the long-term carriage of pneumococci from the nasopharynx. Furthermore, carriage of pneumococci decreases with age, owing to more efficient clearance of pneumococci on subsequent exposure (Lipsitch et al., 2012, van Rossum et al., 2005, Paterson and Mitchell, 2006). Consequently, immune clearance may play an important role in clearing streptococci from the nasopharynx with age. The pneumococci may further be better adapted than other streptococci to colonise this sites, which is indicated by its higher rates of colonisation in infants compared to the viridans streptococci (Kononen et al., 2002). As such, children may present an important sub-population for the development of beta-lactam resistance. Children (≤5yrs) are already considered an important source of drug resistant pneumococci,

owing to the high frequency of pneumococcal carriage that occurs among this subset of the population (Bogaert et al., 2004). However, co-colonisation between pneumococci and drug resistance *S. mitis* and *S. oralis* could also occur more frequently among this subset of the population. The failure to detect these co-colonisation events in this study could therefore represent a sample bias. The VacServ inclusion criteria specifically targeted individuals over the age of 15. Whilst children were recruited into the H1N1 study, samples were chosen at random for inclusion into this study. Consequently, it appears that targeting samples taken from the first 2 years of life may offer a more effective method for identifying co-occurrences between pneumococci and other streptococci. It is also possible that the co-occurrence of *S. mitis, S. oralis* and pneumococci occurs as a result of heavy colonisation, or disruption to the normal microbiota as a result of, for example, illness. Increased nasopharyngeal colonisation has previously been identified as a prerequisite for pneumococcal disease (Albrich et al., 2014, Bogaert et al., 2004). Increased colonisation of the nasopharynx by pneumococci can occur as a result of poor mucosal immunity (Ghaffar et al., 1999, Harabuchi et al., 1994, Bogaert et al., 2004). Such increases in pneumococci in the nasopharynx could facilitate their transmission to other sites of the upper respiratory tract. Consequently, recombination between pneumococci and *S. mitis* and *S. oralis* could occur outside of the nasopharynx. Antibiotics, and vaccines that clear bacteria from the nasopharynx (Bogaert et al., 2004) could also be facilitating the transmission of resistance. Under this scenario, clearance of bacteria from the nasopharynx could lead to re-colonisation with other species (Bogaert et al., 2004), or affect the competitive balance between the species normally present (Veenhoven et al., 2003, Bogaert et al., 2004). This could perhaps facilitate colonisation with species such as *S. mitis* and *S. oralis,* which are normally absent from this site.

A further possibility is that the nasopharynx is not the primary site where recombination occurs. As already mentioned, perhaps heavy pneumococcal colonisation of nasopharynx leads to mixing with *S. mitis* and *S. oralis* inhabiting the oropharynx, immediately behind the nasopharynx. Although these sites appear to have distinct colonisation dynamics normally (Kononen et al., 2002),

this does not indicate that mixing of the microbiota between these two communities cannot occur under certain conditions. Interestingly, a previous study, which carried out large scale microbiological testing of nasopharyngeal samples collected from children and adults similarly failed to identify *S. mitis* and *S. oralis* (pers. comm. Bentley 2012). This finding would similarly support the conclusions as described here, that either the opportunities for genetic exchange between *S. mitis, S. oralis* and the pneumococcus occur rarely in the nasopharynx, or that the oropharynx is the more likely site for such exchanges to occur.

4.3.1. Sampling biases

The *in vitro* analysis highlighted a number of important considerations when trying to assess the carriage dynamics of oral bacteria. Mitis Salivarius Agar, in conjunction with blood agar plates appeared to be important for the positive identification of oral streptococci, due to plates becoming over-grown with co-colonising bacteria otherwise. In addition conducting a PCR on the raw sample may be necessary in order to identify whether the species of interest is present prior to culture. The age of the patient from which the nasopharyngeal sample is taken similarly appears to be an important consideration. As mentioned, the VacServ study specifically targeted adults (≥15yrs), who may not carry *S. mitis* and *S. oralis* to the same extent as children. Whilst the Flu surveillance samples should have avoided this bias, as adults and children were both sampled, only a minority of these samples are likely to have been taken from infants, and so targeting of these samples may offer a more effective sampling strategy. An additional issue was the effect of competition between species present in the nasopharyngeal samples. The STGG samples obtained from the Flu surveillance study underwent an overnight incubation period specifically to improve the recovery of pneumococci from these samples. Consequently, an effect of resource competition could have affected the recovery of species from these samples subsequently. Whilst the UTM samples should have provided an unbiased assessment of the species present in the original sample, far fewer of these samples were tested compared to the STGG samples. Crucially, no UTM samples from infants were tested. Therefore it is possible that *S. mitis* and *S. oralis* could

have been detected if UTM samples were widely screened for these species using a PCR approach, and that samples from infants alone were targeted. This approach however could only be used to indicate the presence of these species, but does not indicate whether the bacterial strains could have been isolated from these samples necessary for further *in vitro* characterisation.

The *in silico* analysis indicated that conserved SNP patterns could not be detected for the identification of oral streptococci to the species level. The SNP markers used here were identified as being widely present among the Malawian 16S reads studied, and also across the phylogeny constructed from these reads. Based on sequence similarity using BLAST, these SNP markers were found across several genera. Consequently, these markers were not found to be unique to streptococci, or to species within this family. This indicated that the 16S region amplified in this study was insufficient to distinguish between reads to a species resolution. As such, whilst SNP markers may exist elsewhere in the 16S gene that can be used for species identification, the region amplified in this study was not sufficient for this purpose. Instead, SNP divergence generally appeared to more effectively distinguish between taxa, lacking any other method to distinguish between the reads in this study. This approach, based on read similarity, indicated that only the six taxa closest to the reference pneumococcal 16S read appeared to belong to the streptococcal class. In addition, only one of these reads appeared to associate more strongly with a non-pneumococcal streptococcus in the BLAST database. Clusters of divergent SNPs did occur in a further three stains. This could suggest they were atypical pneumococci, however, as the raw samples were no longer available, it was not possible to subsequently test the true identity of these strains *in vitro*. Based on the BLAST similarity results however, there was little support that *S. mitis* or *S. oralis* could be identified across the 16S reads analysed here. The failure to detect these species from the Malawian 16S was consistent with *S. mitis* and *S. oralis* not being present, but the detection method developed was equally untested *in vitro*.

The 16S dataset used was compiled from the V3-V5 region. Whilst this region allows for longer continuous reads, which can aid subsequent classification,

considerable debate exists with regards to which primer sets should be used in 16S studies (Wahl et al., Chakravorty et al., 2007). Interestingly Chakravorty and colleagues (2007) suggest that the V4 and V5 regions, along with V7 and V8, offer poor targets for classification of genus, or species. Consequently, other regions of the 16S gene may offer better targets for identification of the streptococcal group. Whilst this may be true, even commercial kits that have been developed for the identification of pneumococci, such as AccuProbe, appear to struggle to distinguish between the oral streptococci. Although AccuProbe was initially claimed to offer 97.9% sensitivity, and a specificity of 100% when compared to conventional methods of pneumococcal identification (Lindholm and Sarkkinen, 2004), the closely related and recently discovered *S. pseudopneumoniae* tests positive for pneumococci using AccuProbe (Arbique et al., 2004). Consequently, it appears that the 16S gene may be a poor target for the accurate differentiation between this group generally. Whilst this suggest the *S. mitis* and *S. oralis* primers developed by Park and colleagues 2012 and 2010 respectively may offer a solution to the differentiation between the mitis group, additional *in vitro* characterisation still appears necessary in order to confirm these findings.

In conclusion, the co-colonisation between *S. pneumoniae* and either *S. mitis* or *S. oralis* could not be identified using both *in vitro* and *in silico* methods. A number of caveats did exist with regards to the samples available for study, which have been discussed in detail above. However, the results presented here indicate that the co-carriage of these species in the nasopharynx occurs infrequently at the population level. This suggests either co-colonisation and genetic exchange is limited to a particular subset of the population, such as infants, or that such occurrences may happen outside of the nasopharynx.

5 Gene Mosaicism and Beta-lactam Resistance in Malawi

5.1 Introduction

Penicillin and ceftriaxone currently form the first line defences against pneumococcal diseases in Malawi (Cornick et al., 2011). Globally, resistance to these two antimicrobials has occurred widely, which can be attributed to their extensive usage, particularly in the developed world (Albrich et al., 2004, Baquero et al., 2002, Hjalmarsdottir and Kristinsson, 2014). The widespread occurrence of resistance has been facilitated by the success of pneumococcal clones that have acquired MDR, before spreading globally (section 1.3.8)(Munoz et al., 1991, Sa-Leao et al., 2002). For developing countries such as Malawi, with limited medical resources, the spread of MDR clones is a major epidemiological concern. Penicillin resistance, based on meningitis resistance breakpoint, is already relatively widespread in Malawi, occurring in close to 20% of pneumococci isolated from IPD at QECH (Everett et al., 2011). The first cases of ceftriaxone resistance have also recently been identified, which follows a gradual loss in pneumococcal susceptibility to this antimicrobial since 2003, when ceftriaxone first became available in the country (pers comm. Everett 2015). A recent genomic analysis has also indicated that several PMEN (section 1.3.8) lineages are now circulating in Malawi, demonstrating the ability for pneumococci to spread on a global scale, and potential for MDR clones to establish in Malawi (Everett et al., 2012). However PMEN27 (ST217), the most prevalent of these clones, whilst associated with virulent disease, is rarely MDR. In addition, although MDR clones are expected to become more prevalent in the future, drug resistance has not been identified as being restricted to any one-sequence type in Malawi (Everett et al., 2012).

Beta-lactam resistance is frequently associated with *pbp* modification, although additional genes are also involved in resistance (Lloyd et al., 2008, Mascher et al., 2006). Three of the six pneumococcal *pbp* alleles frequently appear highly divergent at the nucleotide level when compared to susceptible forms. Such divergence has been attributed to gene mosaicism, which defines the interspersed occurrence of sections of genetic sequence throughout a gene,

which are derived from different ancestries. Such divergent sequence blocks may be derived through recombination within, and from other species (Hollingshead et al., 2000, Lorenz and Wackernagel, 1994). Mosaicism has been widely documented in resistant forms of pneumococcal *pbp1a*, *pbp2b* and *pbp2x* genes (Smith and Klugman, 1998, Munoz et al., 1992) (section 1.5). Recombined sequence blocks were found to differ by up to 21% in one study, indicating recombination with other species as being an important process in the occurrence in beta-lactam resistance among pneumococci (Dowson et al., 1993). Recombination with streptococcal co-colonisers of the nasopharynx has subsequently been hypothesised as an important source of beta-lactam resistance, with particular focus falling on *S. mitis* and *S. oralis* (Dowson et al., 1993). These two species are closely related to the pneumococcus based on their sequence divergence. This appears to increase the efficiency with which pneumococci can undergo recombination with these two species relative to other more closely or more distantly related species (Humbert et al., 1995). This phenomenon results from the activities of the pneumococcals mismatch repair, "Hex", system. The Hex system is a DNA replication editor, which removes potentially mutagenic nucleotide mismatches (Humbert et al., 1995). The system is found to efficiently remove variation from the least divergent sequences during the donor-recipient heteroduplex stage of transformation. Contrastingly, at greater levels of sequence divergence, there is insufficient similarity between the donor and recipient strains for recombination to take place. However, sequence that occupies an intermediate level of divergence on this scale appears to be efficiently incorporated owing to the Hex system being saturated by mismatches, rendering it unable to perform its function (Humbert et al., 1995). Humbert and colleagues (1995) demonstrated how *S. mitis* and *S. oralis* were sufficiently divergent at the sequence level to saturate the Hex system, but similar enough to successfully recombine with pneumococci. In accordance with this study, shared genetic material within the *pbp* genes of *S. pneumoniae* and these species has been found among beta-lactam resistant pneumococci (Dowson et al., 1993).

Whilst *pbp*-mosaicism and sequence divergence (relative to susceptible isolates) can be indicative of beta-lactam resistance, a loss in beta-lactam susceptibility is typically associated with the incorporation of a limited number of nucleotide polymorphisms that affect the interaction between PBP and substrate at three conserved motifs (Sauvage et al., 2008). Consequently, although *pbp* alleles may contain numerous recombined fragments, variation that affects these motif regions is likely to have the greatest affect on the overall beta-lactam susceptibility (Granger et al., 2005, Dowson et al., 1993). The donor material must also contain resistance-encoding SNPs in order for beta-lactam susceptibility to be affected (Dowson et al., 1993). For this reason, the dissemination of resistance among the oral streptococci is likely to have a profound affect on the ability of pneumococci to acquire this material.

S. mitis and *S. oralis* are thought frequent colonisers of the oral cavities, and are found to be widely beta-lactam resistance (Teng et al., 1998, Mandell et al., 2002, Farber et al., 1983). This has lead to the view that *S. mitis, S. oralis,* as well as other drug resistant pneumococci, have acted as globally circulating pools of resistance encoding genetic material, which can be availed of by susceptible pneumococci (Reichmann et al., 1997, Sauerbier et al., 2012a).

The degree to which such populations have contributed to drug resistance within the Malawian pneumococcal population is currently unknown. However the combination of antibiotic pressure, and vaccination will alter the circulating STs currently present, and could favour the emergence of drug resistant lineages (Everett et al., 2012), or favour the acquisition of resistance encoding material among pneumococci. This project was intended to determine the level of *pbp* variation present among Malawian pneumococci, and determine the likely origin, and contribution of globally circulating drug resistant bacterial populations to beta-lactam resistance within the Malawian pneumococcal population.

5.1.1. Study Aims

- To assess the level of diversity among key beta-lactam resistance genes in Malawi and identify the prevalence of key mutations associated with beta-lactam resistance.
- To determine the contribution of recombination to this diversity.
- To determine the possible origins of these recombined sequences, focussing on the contribution of genetic exchange with *S. mitis* and *S. oralis* and other globally circulating pneumococcal populations.

5.2 Study samples

A collection of 682 whole genome sequenced Malawian pneumococcal isolates formed the basis of this study. These isolates were collected between 2003 and 2011 by medical staff at QECH, Malawi's largest tertiary referral hospital (section 2.1). Poor recovery of pneumococci from samples prior to 2003 meant that it was not possible to sequence isolates from this period. The isolates used in this study were sequenced as part of previous projects that have assessed the genetic diversity of pneumococci prior to PCV13 introduction (Everett et al., 2012), and investigated the association between invasive pneumococcal disease and genotype (Kulohoma unpublished 2015). As such, a mix of samples collected from cases of asymptomatic carriage (~20%) and invasive pneumococcal disease (~80%), and from children and adults were present. Metadata was available for approximately half of the study collection, and included penicillin disc susceptibility information, which is collected routinely for clinical pneumococcal samples by QECH staff. *In silico* serotype detection (section 2.7.4) revealed that at least 48 serotypes were present. However, due to the importance of serotype 1 pneumococcal disease in Malawi, this serotype was over-represented in the dataset, comprising ~20% of sequences. In total 93 STs were identified, the most common of these was ST217 (78 isolates), a serotype 1 lineage, belonging to PMEN27 clone (Everett et al., 2012). This project was primarily focused with identifying nucleotide polymorphism present within key beta-lactam resistance genes among the Malawian pneumococci population. Consequently, although certain pneumococcal serotypes were over-represented

in the dataset, these sampling biases were not thought to effect the overall detection of resistant alleles present. However, it should be noted that despite serotype 1 being a considerable cause of invasive disease across Africa, it is rarely associated with drug resistance. Consequently, it was important to consider this collection bias when inferring about the pneumococcal population in Malawi as a whole.

5.2.1 Beta-lactams in Malawi

Ceftriaxone and penicillin are used preferentially for treating meningitis and pneumonia at QECH, Malawi (Table 24). Ceftriaxone was introduced into Malawi in 2004, where it was initially used in restricted cases as a frontline antibiotic. However, since 2007 it's usage was increased to include sepsis and meningitis. Everett and colleagues (2011) recorded the first reductions in susceptibility to ceftriaxone in 2009, from an MIC of 0.016μg/mL to 0.125μg/mL, and the first cases of clinical resistance have now been found (pers. comm. Everett 2015). In contrast, penicillin has been used widely over the duration of this study period, which is reflected by higher resistance levels (Everett et al., 2011).

Antibiotic	Treatment usage in Malawi	PBP affinities
Penicillin	Pneumonia	2x, 2b, 1a (Barcus, 1995)
Ceftriaxone	Pneumonia and meningitis	2x, 1a (Muñoz, 1992)

Table 24: The usage of penicillin and ceftriaxone in Malawi at current. Target PBP enzymes are also indicated.

Penicillin susceptibility information was available for 320 isolates within the Malawian sample collection. Using meningitis breakpoints (CLSI, 2015), of the 320 isolates, 17% were classed as penicillin resistant using meningitis breakpoints, similar to published figures based on pneumococcal surveillance in Malawi (Everett et al., 2011, Cornick et al., 2011). Ceftriaxone susceptibility testing is not undertaken routinely at QECH. Consequently Etest®-ing according to BSAC guidelines was undertaken on 24 isolates (BSAC, 2013). No ceftriaxone resistance was detected, with most isolates highly susceptible (MIC of <0.016μg/mL). Two isolates however were found to have MICs of 0.125μg/mL,

consistent with the view that susceptibility to ceftriaxone is beginning to reduce (Everett et al., 2011).

Given the reduced susceptibilities to these two beta-lactams, substantial modifications within the genes *pbp2x*, *pbp2b* and *pbp1a* would be expected to be present. Consequently, diversity across these *pbp* genes was assessed.

5.2.2 Sequence Assembly

Short reads were assembled against the reference *S. pneumoniae* ATCC700669 genome using SMALT (section 2.7.2). Although divergent from the Malawian isolates, this sequence was chosen, as overall read coverage among the Malawian sequences was still high (>90% of reads mapped) and due to the high level of annotation published previously for this isolate (Croucher et al., 2009). The reference annotation was used to identify the start and end positions of *pbp* alleles, and other genes of interest in the study population. Aligned alleles were extracted and a maximum likelihood phylogenetic tree constructed for each gene. Due to variation in the mapping quality, isolates were excluded from the analysis in cases where less than 50% of the gene sequence had mapped to the reference (Figure 46).

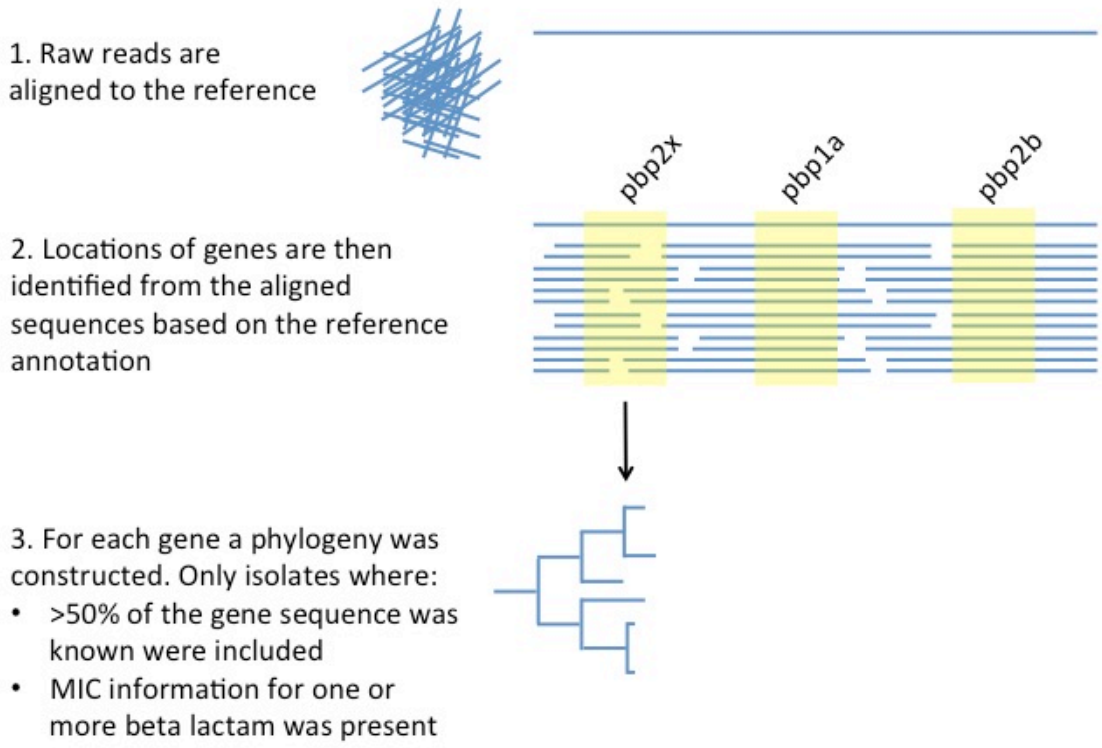


Figure 46: Flow diagram to show how isolates were assembled and phylogenies constructed. Short reads are first aligned to a reference (1) producing an alignment file. Based on the reference annotation, the position of key genes can then be identified, and used to construct a phylogeny (3).

Isolates for which no penicillin susceptibility information was available were also excluded. This allowed tree divergence patterns to be compared directly to the susceptibility information.

For each *pbp* gene of interest, the phylogeny was plotted alongside the penicillin and ceftriaxone susceptibility information. Phylogenies were rooted based on the beta-lactam susceptible R6 strain, to aid comparison. Below each phylogeny, the scale bar was plotted, which indicates the mean number of nucleotide changes occurring per site relative to branch length (Stamatakis et al., 2005). Penicillin susceptibility was based on meningitis breakpoints, whereas, for ceftriaxone the raw MIC information was available. So that the SNPs contributing to the branching pattern of the phylogeny could be viewed, and to visualise how these SNPs were distributed within the gene, these were plotted alongside each isolate relative to their position within the gene (Figure 47, Figure 49, Figure 51).

5.2.3 *Pbp2x*

PBP2x is thought to be the primary target for penicillins and cephalosporins, and is often heavily altered in beta-lactam resistant pneumococci (Sibold et al., 1994, Chi et al., 2007). The phylogenetic tree for *pbp2x* (Figure 47) was consistent with this view. Sequence homology was estimated by counting the number of unique SNPs, or amino acid alterations, present in study isolates relative to the control for each gene of interest. The level of sequence homology between study isolates and the susceptible R6 reference gene was estimated at between 88.79% and 99.93% (98.02% arithmetic mean). Alleles were estimated to contain on average 34.16 SNPs relative to the 2253bp *pbp2x* reference gene. However, the median branch length corresponded to a value of 0.02 (2dp), with the longest branches an order of magnitude larger than this (0.82), suggesting that variation was highly skewed among the study isolates. This was supported by the amino acid translation. Here some isolates were found to share 100% homology to the R6 genome, whereas this value was reduced to 92.20% in the most divergent alleles. This was also apparent from the phylogeny, where the majority of isolates are found to occur at basal position in the tree, with a limited number of highly divergent clusters. There was no observable trend between branch length and ST or serotype in these outlying clusters. This indicates that those isolates with reduced beta-lactam susceptibilities do not belong to the same pneumococcal lineages.

Furthermore, the high SNP density apparent within a limited number of isolates was inconsistent with their having been introduced through the gradual accumulation of mutations, at a constant rate between isolates. Under this scenario a more random and equally distributed accumulation of mutations across these alleles would be expected (Kimura, 1983). Consequently, there appeared to be a strong indication for localised recombination within particular clusters of isolates. This is further reflected when viewing the distribution of SNPs within each allele (Figure 47). Here, large blocks of SNP dense regions are apparent within the more divergent alleles. These blocks were not restricted to the locations of the key motifs within the gene however, indicating apparently random recombination events had occurred throughout this gene.

Variation was not found to be unique to the resistant isolates, but occurred among both susceptible and resistant isolates alike. This could be an artefact of penicillin susceptibility information represented by breakpoints, rather than as a gradation to resistance. Furthermore, the meningitis breakpoint used to differentiate susceptible from non-susceptible strains here, represents the point where resistance requires additional *pbp1a* modification (Smith and Klugman, 1998). As *pbp2x* modification is only required during the initial loss of penicillin susceptibility, this is perhaps why a stronger association between susceptibility and nucleotide polymorphism was not found. However, given that several SNP clusters occur outside of the three conserved motif regions, it appears that much of the variation within these alleles is unlikely to have resulted in beta-lactam susceptibility changes.

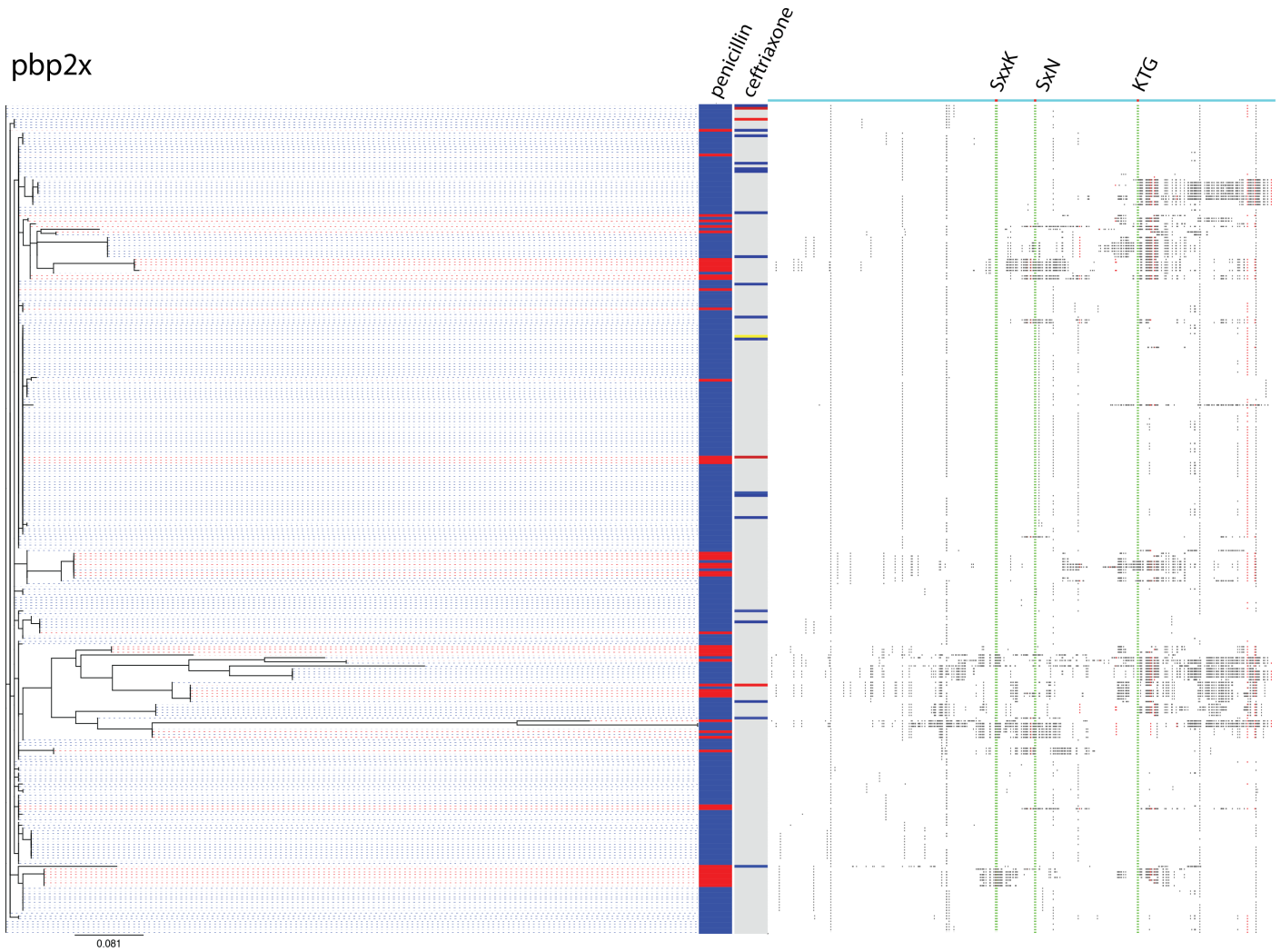


Figure 47: A phylogenetic tree constructed for the Malawian isolates based on the nucleotide variation present in the *pbp2x* gene. Penicillin and ceftriaxone susceptibility information is indicated (blue= susceptible, red= resistant). SNPs are plotted for each isolate relative to their positions in the gene. Conserved motifs are also indicated.

To investigate the relationship between branch length and penicillin susceptibility further, root to tip distances for each isolate were plotted, ranked from smallest to largest (Figure 48). In this way a direct assessment of SNP divergence in the tree relative to penicillin susceptibility could be determined (Figure 48).

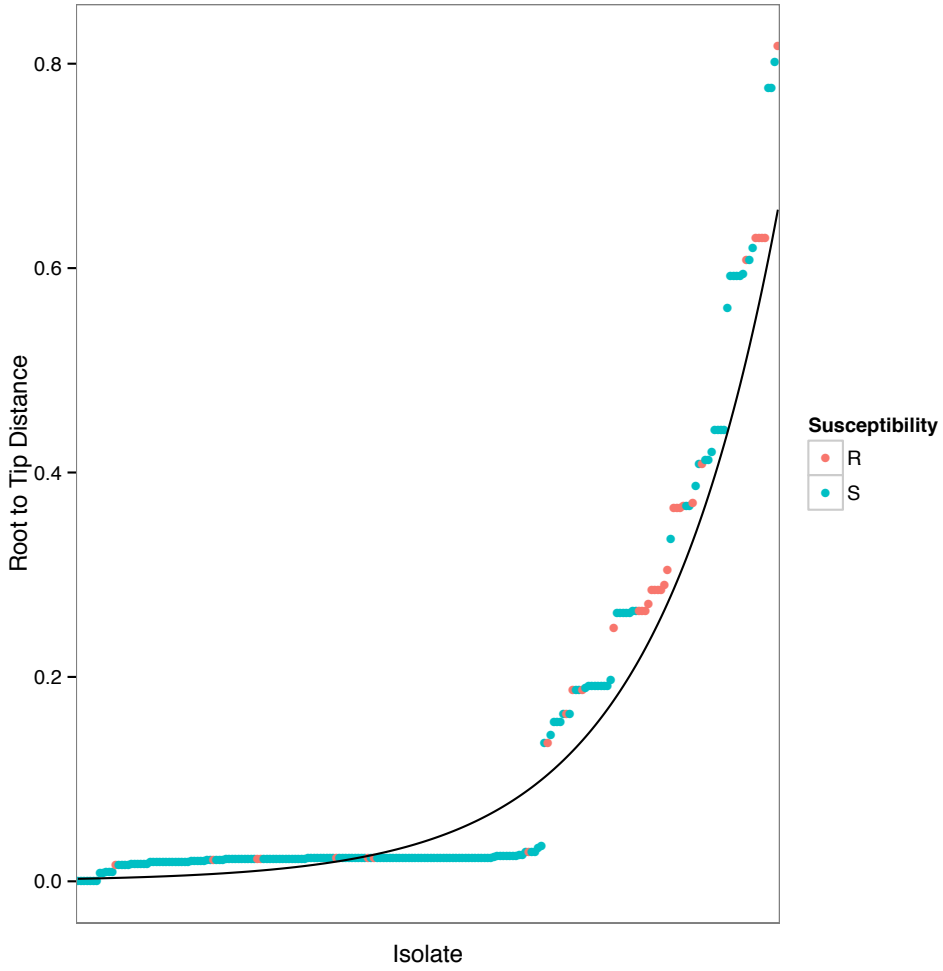


Figure 48: Root to tip length for each isolate. R: penicillin resistant, S: penicillin susceptible according to meningitis breakpoints. An exponential curve was plotted, as an approximation of the true distribution.

As shown in Figure 48, the majority of resistant isolates (red points) occurred along more divergent branches of the tree. However, several isolates occurring at the base of the tree were also found to carry penicillin resistance. This suggests that branch length alone is not a reliable predictor of beta-lactam susceptibility. The graph also indicated that nucleotide divergence appears to follow an exponential distribution (R^2=0.57, $p<2.2x10^{-16}$) within the *pbp2x* gene. As was apparent from the phylogeny, this distribution shows that most isolates carried a *pbp2x* allele that was highly similar to the susceptible R6 type gene. However, a

point is reached, where there occurs a sudden rapid increase in branch length, which appears to rise without plateauing. This sudden divergence is indicative of recombination events introducing large numbers of SNPs within a single event (Chewapreecha et al., 2014a). The sudden exponential rise could also be indicative of a biological feature of the pneumococcus. The Hex mis-match repair system of pneumococcus has been described to efficiently remove highly related sequences, whereas recombining sequences of intermediate levels of divergence (<17.5%) are found to overload the system (Humbert et al., 1995). This saturation of the hex system means that sequence transformation efficiency occurs with a linear frequency relative to rate of divergence (Humbert et al., 1995, Claverys et al., 2000). Eventually the donor sequence becomes too divergent to effectively recombine, hence the rapid decline in transformation efficiency Humbert and colleagues (1995) identified after approximately 17.5% sequence divergence.

The translated amino acid sequence was also assessed in order to identify amino acid polymorphism present within the *pbp2x* gene. Conserved amino acid changes are widely recorded among MDR resistant clones, therefore this investigation was restricted to those amino acid changes that have a proven relationship to changes in beta-lactam susceptibilities. A Fisher's Exact test was used to determine whether the presence or absence of these polymorphisms was statistically associated with penicillin susceptibility. Table 25 summarises the resistance associated SNPs identified. This included the amino acid alteration T338A within the SxxK motif (Zerfass et al., 2009), which was present in 7 of 250 isolates tested. A further 8 amino acid changes were identified within the dataset (Table 25). Interestingly two of these were highly conserved across the isolates, Q552E and D567N, being found in 72 and 70% of isolates tested respectively. Both of these mutations have been found to correspond with changes in beta-lactam susceptibility (Hakenbeck et al., 1999, Carapito et al., 2006), and furthermore, both were significantly associated with penicillin resistance (Q552E, p= 0.045, and D567N, p= 0.0125). However, due to the suspected sampling biases present in the initial dataset, it is difficult to infer the prevalence of these resistance mechanisms in the general pneumococcal population in Malawi.

Amino acid change	Absent	Present	Frequency (%)	Association with resistance	SNP confers reduced susceptibility to (class):	Reference
T338A	257	7	3	P=0.0001	Penicillins	(Zerfass et al., 2009)
R384G	34	14	29	P=0.0001	Penicillins and Cephalosporins	(Carapito et al., 2006)
H394Y	117	13	10	P=0.0271	Penicillins and Cephalosporins	(Pernot et al., 2004)
Q552E	23	58	72	P=0.0045	Penicillins and Cephalosporins	(Hakenbeck et al., 1999)
S389L	46	22	32	P=0.0007	Penicillins and Cephalosporins	(Dessen et al., 2001)
N514H	127	11	8	P=1	Penicillins and Cephalosporins	(Dessen et al., 2001)
V516I	84	9	10	P=0.1525	Penicillin and cefotaxime	(Pernot et al., 2004)
D567N	28	64	70	P=0.0125	Penicillins and Cephalosporins	(Carapito et al., 2006)
N605T	201	4	2	P=1	Penicillins and Cephalosporins	(Sadowy et al., 2010a)

Table 25: Amino acid polymorphisms identified in the literature as associated with beta-lactam resistance, and present within the Malawian *pbp2x* gene. The SNP's prevalence is also indicated (for isolates in which MIC information is available). In addition the p-value corresponds to a Fisher's Exact test, determining whether an association with reduced penicillin was identified in the Malawian isolates with that SNP.

5.2.4 *Pbp2b*

The phylogeny of *pbp2b* was similar in structure to that of *pbp2x*, with distinct clusters of highly divergent alleles occurring throughout the tree (Figure 49).

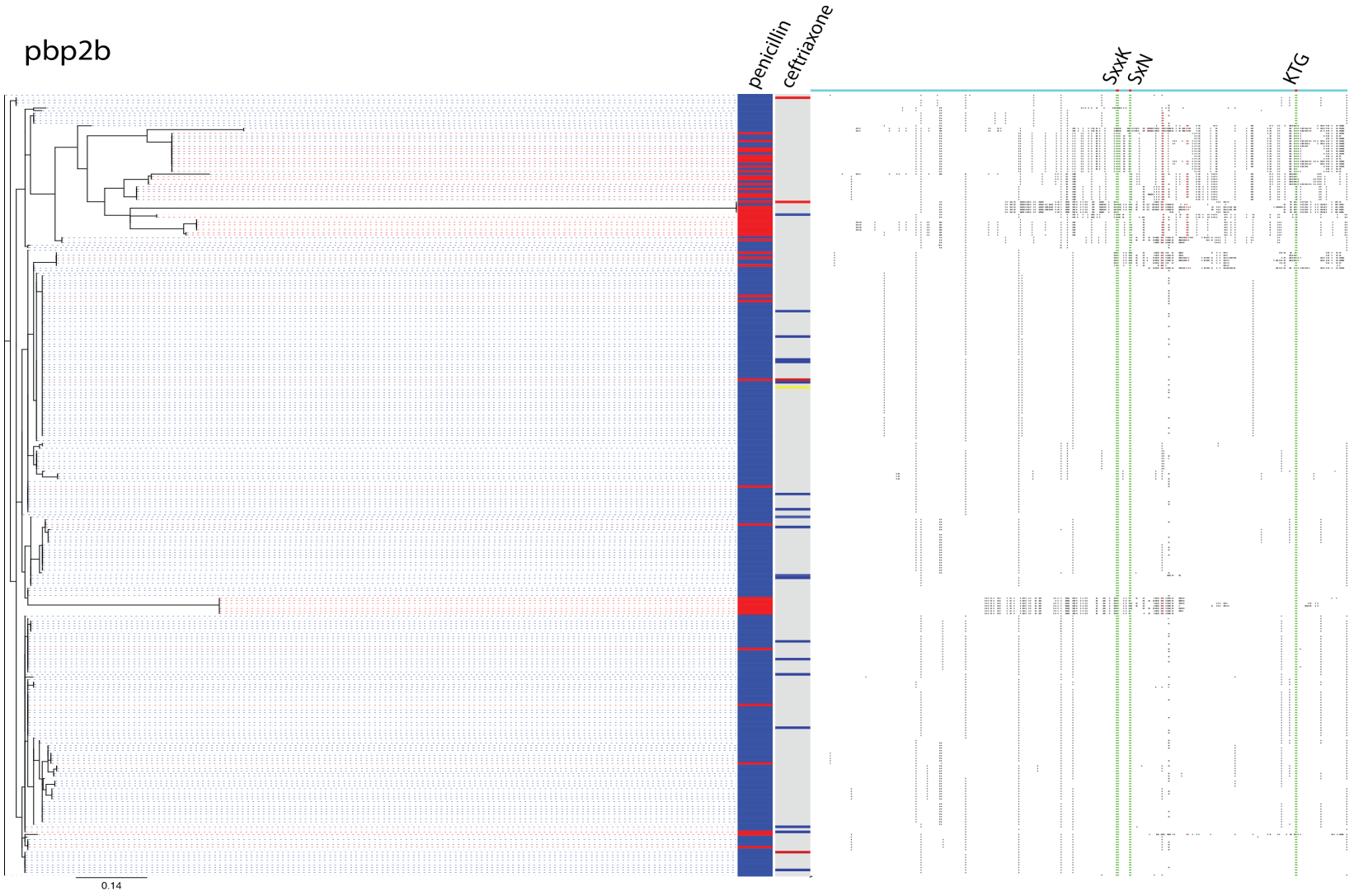


Figure 49: A phylogenetic tree constructed for the Malawian isolates based on the nucleotide variation present in the *pbp2b* gene. Penicillin and ceftriaxone susceptibility information is indicated (blue= susceptible, red= resistant). SNPs are plotted for each isolate relative to their positions in the gene. Conserved motifs are also indicated.

Sequence homology was calculated as described above, and was slightly higher than identified in *pbp2x*, ranging from between 90.08% and 99.95% (98.79% arithmetic mean). The average number of SNPs per isolate was also reduced to 22.38 as a result. Similarly the amino acid sequence shared a greater homology to the reference strains than pbp2x (99.41% arithmetic mean). Large regions of high SNP density were found clustered among particular isolates similar to what was observed in *pbp2x*, and was indicative of recombination. There also appeared to be a greater association between branch length and penicillin resistance. However, plotting the branch length for each isolate, resistant strains can similarly be seen to occur at all branch lengths. This again suggests that branch length, and by inference tree divergence patterns alone were poor predictors of penicillin susceptibility. Branch lengths showed a similar exponential relationship (R^2=0.69, $p< 2.2x10^{-16}$) as had been observed in *pbp2x*. The fitted curve was actually found to under fit the data, indicative of a greater number of alleles sharing high homologies with the susceptible R6 reference, with a sharper increase in isolate divergence (Figure 50).

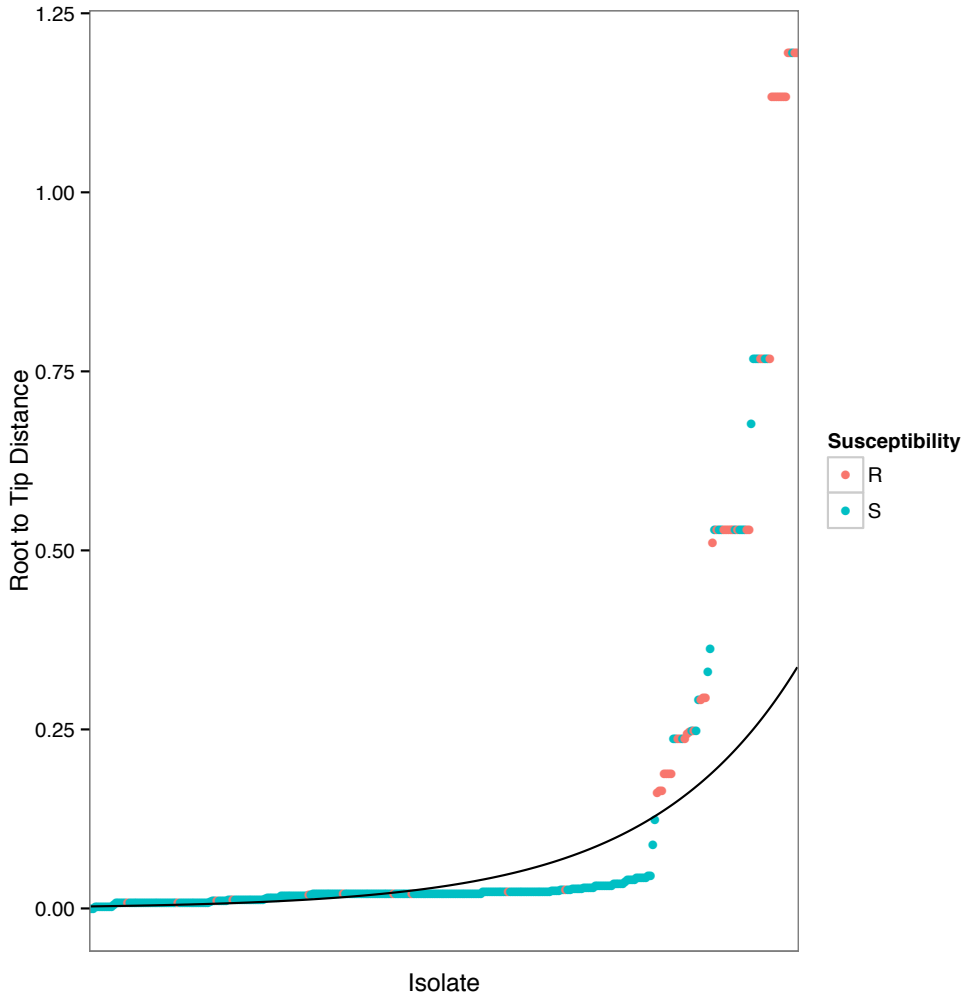


Figure 50: Plot of root to tip length for each isolate, ranked in order, smallest to longest (least, to highest SNP divergence). R= penicillin resistant, S= penicillin susceptible according to meningitis breakpoints. An exponential curve was plotted, as an approximation of the true distribution. The curve is found to fit the data poorly, with a limited number of isolates found to have very long branch lengths and by inference, carry most of the nucleotide diversity present in this gene.

Comparison with the literature identified three amino acid alterations that have either been proven, or suggested to have been involved in beta-lactam resistance (Table 26). The most prevalent of these was the T446A mutation, found in 22% of isolates. Interestingly, this mutation is found to reduce penicillin affinities by over 60% among susceptible isolates (Pagliero et al., 2004). Although the T426K mutation was also identified, this was not found to correlate with penicillin resistance (p=1). Given the high association between this amino acid substitution and penicillin resistance in the literature (Davies et al., 2010), it appears that either complementary mutations are necessary for resistance to result, or that the limited susceptibility information recorded did not allow any relationship to be identified in the present analysis.

Amino acid change	Absent	Present	Frequency (%)	Association with resistance	SNP confers reduced susceptibility to (class):	Reference
T426K	105	4	2	P=1	Penicillins and Cephalosporins	(Davies et al., 2010)
T446A	232	65	22	P=0.0001	Penicillins	(Pagliero et al., 2004)
E476G	241	34	12	P=0.0001	Penicillins	(Pagliero et al., 2004)

Table 26: Amino acid polymorphisms identified in the literature as associated with beta-lactam resistance, and present within the Malawian *pbp2b* gene. The SNP's prevalence is also indicated (for isolates in which MIC information is available). In addition the p-value corresponds to a Fisher's Exact test, determining whether an association with reduced penicillin was identified in the Malawian isolates with that SNP.

5.2.5 *Pbp1a*

In terms of divergence from the R6 reference gene, variation among *pbp1a* alleles was more similar to that of *pbp2x*. Homology was estimated to lie between 89.32% and 99.86% (98.52% arithmetic mean), relative to *pbp2x* 88.79% and 99.93% (98.02% arithmetic mean), compared to the reference. Fewer SNPs however were introduced on average in each isolate (23.87), when compared to *pbp2x* (34.16). The phylogenetic tree further suggested that alleles were generally less conserved, with branch lengths generally appearing longer (Figure 51). In addition some association between branch length and penicillin susceptibility was identified. In order to investigate this further, branch lengths were plotted in order of increasing size, relative to the susceptible R6 strain (Figure 52). An exponential curve was fitted (R^2=0.66, $p<2.2x10^{-16}$) in order to compare with similar plots for *pbp2x* and *pbp2b*. However, whilst the curve fits the data well initially, distinct from *pbp2x* and *pbp2b* diversity appears to plateau, at a low level. Interestingly, there was also a clearer association between penicillin susceptibility and *pbp1a* divergence. Mutations in *pbp1a* have been shown to be required to confer penicillin resistances of above 0.25µg/mL (Smith and Klugman, 1998). Given that the meningitis breakpoint for penicillin is 0.5µg/mL it might be expected that a greater correlation occurred between branch length and penicillin susceptibility for these alleles. However, once again, penicillin resistance was not restricted to the more divergent alleles.

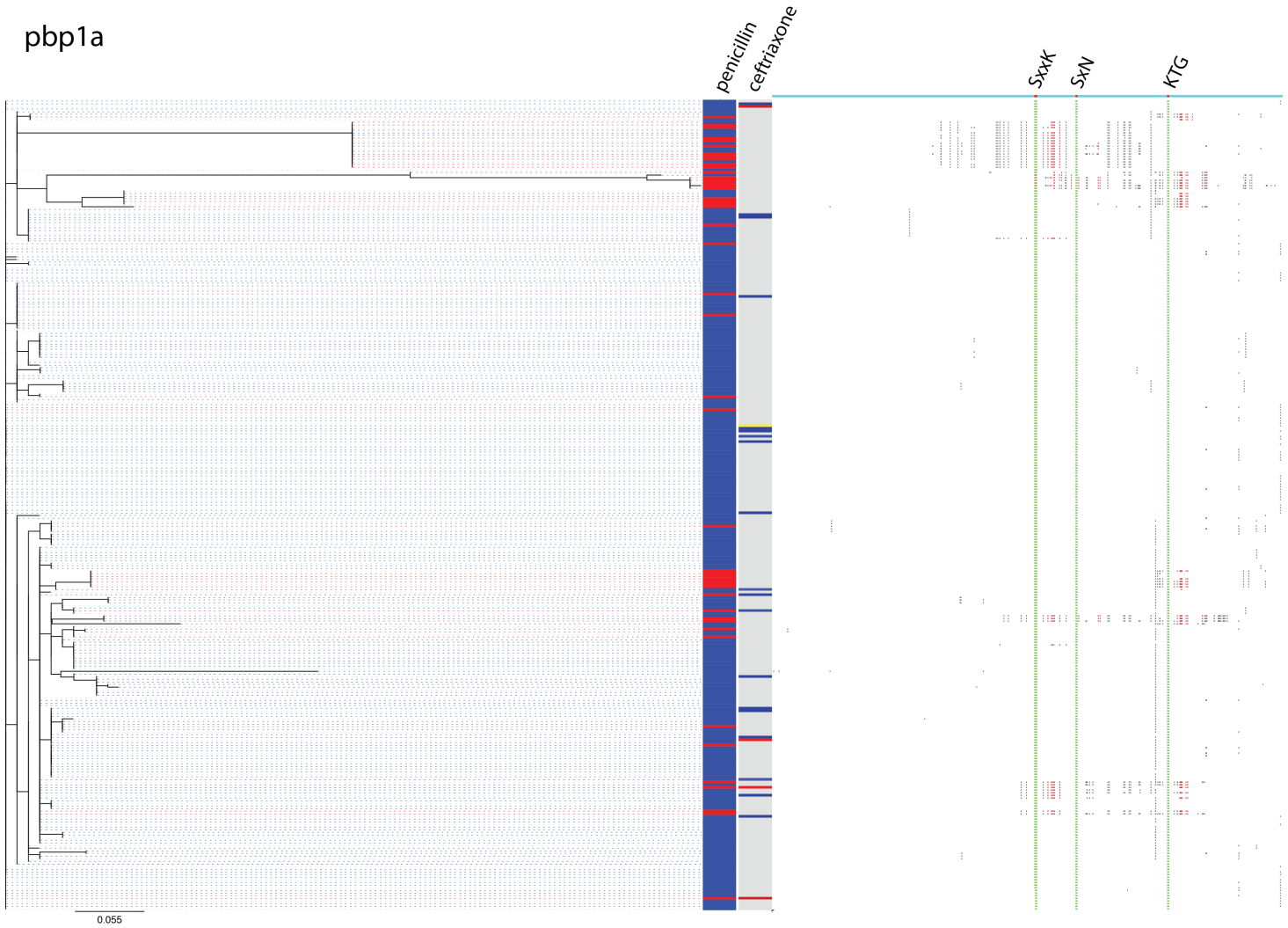


Figure 51: A phylogenetic tree constructed for the Malawian isolates based on the nucleotide variation present in the *pbp1a* gene. Penicillin and ceftriaxone susceptibility information is indicated (blue= susceptible, red= resistant). SNPs are plotted for each isolate relative to their positions in the gene. Conserved motifs are also indicated.

Investigating the isolates present at the plateau region, two serotypes were abundant, forming two clusters. Penicillin resistance was identified in 9 of the 13 isolates in the serotype 14 cluster, all belonging to ST63. In the second cluster, 4 of 7 isolates carried penicillin resistance, although these belonged to various STs. No distinct amino acid changes were found to be common to resistant isolates relative to the susceptible isolates present in each of these divergent groups. *Pbp1a* alleles were however highly conserved within each cluster and commonly carried amino acid substitutions that had previously identified as being associated with penicillin resistance. Consequently, the variation in susceptibility within these divergent clusters, could reflect the presence or absence of addition nucleotide polymorphisms within *pbp2x* and *pbp2b*.

The departure from the exponential curve therefore appears to have arisen through a clonal expansion of *pbp1a* alleles associated with the lineages in this plateau region. Furthermore, the lack of divergence within *pbp1a* relative to *pbp2x* and *pbp2b* could suggest a lack of selective pressure for greater divergence.

Pbp1a was the most divergent of the *pbp* alleles in terms of amino acid sequence, ranging between 92.30% and 99.82% homology with the reference sequence. This suggests that although fewer SNPs were present within *pbp1a* alleles, a greater number of these conferred amino acid changes than in the previously investigated *pbps*.

In addition, a number of these amino acid changes statistically associated with penicillin resistance (Table 28). However the T371A substitution described by Smith and Klugman (1998) was an exception to this trend (p=0.0629), again indicating complementary *pbp* mutations are necessary for beta-lactam resistance. Interestingly, these mutations were present in approximately 16% of cases, similar to the 17% penicillin resistance level among the isolates.

Lacking ceftriaxone MIC information, little inference could be made about how ceftriaxone MICs compared to penicillin MICs, although no association was

obvious. Additional ceftriaxone susceptibility testing was not possible as the original isolates were no longer available. The analysis however suggests that routine ceftriaxone susceptibility testing may be an important future consideration in order to identify how resistance emerges to this antibiotic.

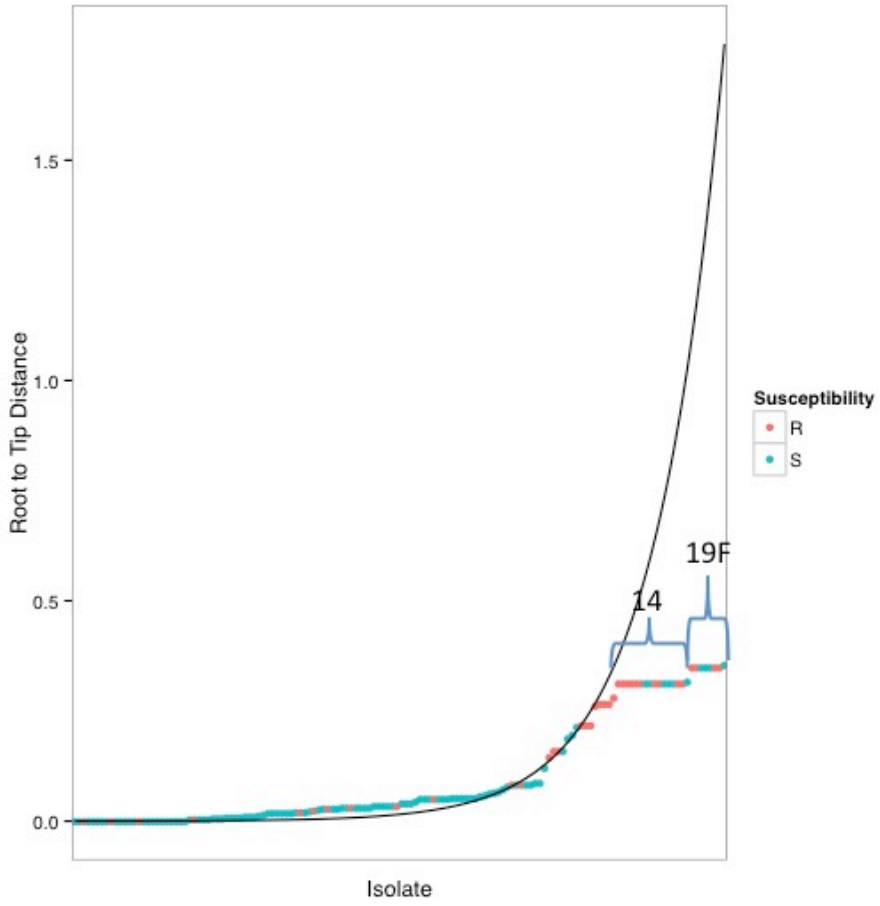


Figure 52: Plot of root to tip length for each isolate, ranked in order, smallest to longest (least, to highest SNP divergence). R= penicillin resistant, S= penicillin susceptible according to meningitis breakpoints. An exponential curve was plotted, as an approximation of the true distribution. Two clusters, containing predominantly serotypes 14 (ST63), and 19F isolates are indicated.

Amino acid change	Absent	Present	Frequency (%)	Association with resistance	SNP confers reduced susceptibility to (class):	Reference
T371A	28	5	15	P=0.0629	Penicillins	(Smith and Klugman, 1998)
P432T	12	8	40	P=0.0003	Penicillins	(Granger et al., 2005)
I459M	69	9	12	P=0.0001	Penicillins	(du Plessis et al., 1999)
S462A	73	8	10	P=0.0003	Penicillins	(du Plessis et al., 1999)
T574N	173	30	15	P=0.0001	Penicillins	(Smith and Klugman, 1998)
S575T	171	31	15	P=0.0001	Penicillins	(Smith and Klugman, 1998)
Q576G	173	31	15	P=0.0001	Penicillins	(Smith and Klugman, 1998)
F577Y	172	32	16	P=0.0001	Penicillins	(Smith and Klugman, 1998)
L583M	202	31	13	P=0.0001	Penicillins	(Smith and Klugman, 1998)
A585V	233	31	12	P=0.0001	Penicillins	(Smith and Klugman, 1998)

Table 27: Amino acid polymorphisms identified in the literature as associated with beta-lactam resistance, and present within the Malawian *pbp1a* gene. The SNP's prevalence is also indicated (for isolates in which MIC information is available). In addition the p-value corresponds to a Fisher's Exact test, determining whether an association with reduced penicillin was identified in the Malawian isolates with that SNP.

5.2.6 PBP divergence Summary

Analysis of the *pbp2x*, *pbp2b* and *pbp1a* alleles within the Malawian dataset indicated that divergence was generally high, but limited to a relatively small number of isolates. Little association between branch length and penicillin susceptibility was found. Ceftriaxone resistance was similarly not found uniquely within the most divergent isolates. Plots of SNPs relative to the position within the genes indicated that these form dense regions, apparently randomly within the genes, indicative of recombination events. Given the sensitivity of particular regions of the *pbp* gene to penicillin interactions, it therefore appears that divergence does not correlate well with resistance, because in many cases, the amino acid changes appear to be introduced into less sensitive (in terms of affecting overall susceptibility) regions of the *pbp* gene. Consequently, it appears important to consider whether the amino acid changes present within each gene are identified in the literature as showing association to penicillin resistance. A number of amino acid alterations were identified which were identified in the literature as affecting beta-lactam susceptibility. However, these did not always correlate with the predicted phenotype recorded in those publications. As such, compensatory mutations between *pbp* alleles clearly play an important role in the manifestation of resistance (Smith and Klugman, 1998, Orio et al., 2011).

A distinct pattern within the branching structure of *pbp1a* was also identified relative to *pbp's 2x* and *2b*. There appeared to be less of diversity within this gene relative to the other *pbp* alleles considered. The possible reasons for this distinct pattern are explored in the discussion, but were not tested here.

5.3 Identifying pbp recombination events

The final analysis was designed to investigate the role of recombination in *pbp* divergence and beta-lactam resistance. In order to carry out this assessment, it was necessary to develop a consistent method to identify recombination events. The method adopted was similar to that of Croucher and colleagues (2011, Croucher et al., 2015), which was used to identify genome-wide recombination events. This method relies on detecting regions across a sequence where SNPs

occur at a higher rate than would be expected, relative to a background rate calculated for the sequence. Theoretically, mutation should introduce SNPs at a random uniform rate with time; hence the background rate should reflect this. Where SNPs occur at a level greater than this, they are hypothesised as having been introduced through a recombination event (Croucher et al., 2015).

This analysis was carried out using the entire Malawian sequence dataset, rather than being restricted to those with penicillin MIC information, as was done previously. Sequence diversity in this was expected to be high, as it represented a mix of different pneumococcal populations, rather than samples from within a single clonal lineage. Consequently identifying recombination events on a genome level would not be specific enough to address the contribution of recombinations within key beta-lactam targets. As such recombination detection was limited to *pbp* genes. Given the potential for other *pbp* genes to affect beta-lactam resistance *in vitro* (other than *pbp3*) these were also included. In addition, several genes that had been identified in chapter 3 as carrying polymorphisms associated with changes in beta-lactam susceptibility were also included. Recombination detection was expanded to include these genes in order to identify whether gene mosaicism could be identified within genes other than *pbps*. This could further support a role for these genes in changes in beta-lactam susceptibilities. The additional genes assessed included *ddlA* and it's promoter region, *murF*, *leuS* and *murM*. *MurM* was included as this has previously been found to carry mutations conferring beta-lactam resistance (Lloyd et al., 2008). A summary of the recombination events identified is provided in Table 28.

5.3.1 *Pbp2x*

In total 103 recombination events were identified in the *pbp2x* gene. Of the 3600 SNPs present, 2653 (74%) were identified as being introduced through 103 recombination events. The ratio of the relative likelihood of a SNP being introduced through recombination rather than mutation was estimated to be 2.8 for this gene, lower than previously calculated values using MLST (Feil et al., 2000) or whole genome data (Croucher et al., 2011). Given that *pbp* genes

represent highly recombinogenic regions, it appears that recombinations were in fact under-detected in this analysis. This is likely to be because recombination events were being identified within single loci rather than genome-wide. As such the background "mutation rate" is likely to be artificially high- as in some cases, recombination events can replace the whole *pbp2x* allele (Dowson et al., 1989, Croucher et al., 2012).

Overall, removing recombined regions from the alignment led to a substantial reduction in branch lengths, so that the association between root-to-tip distances and isolation dates was more closely correlated (prior R^2=0.02, p= $7.8x10^{-3}$, post R^2=0.05, p= $1.6x10^{-4}$)(Figure 53).

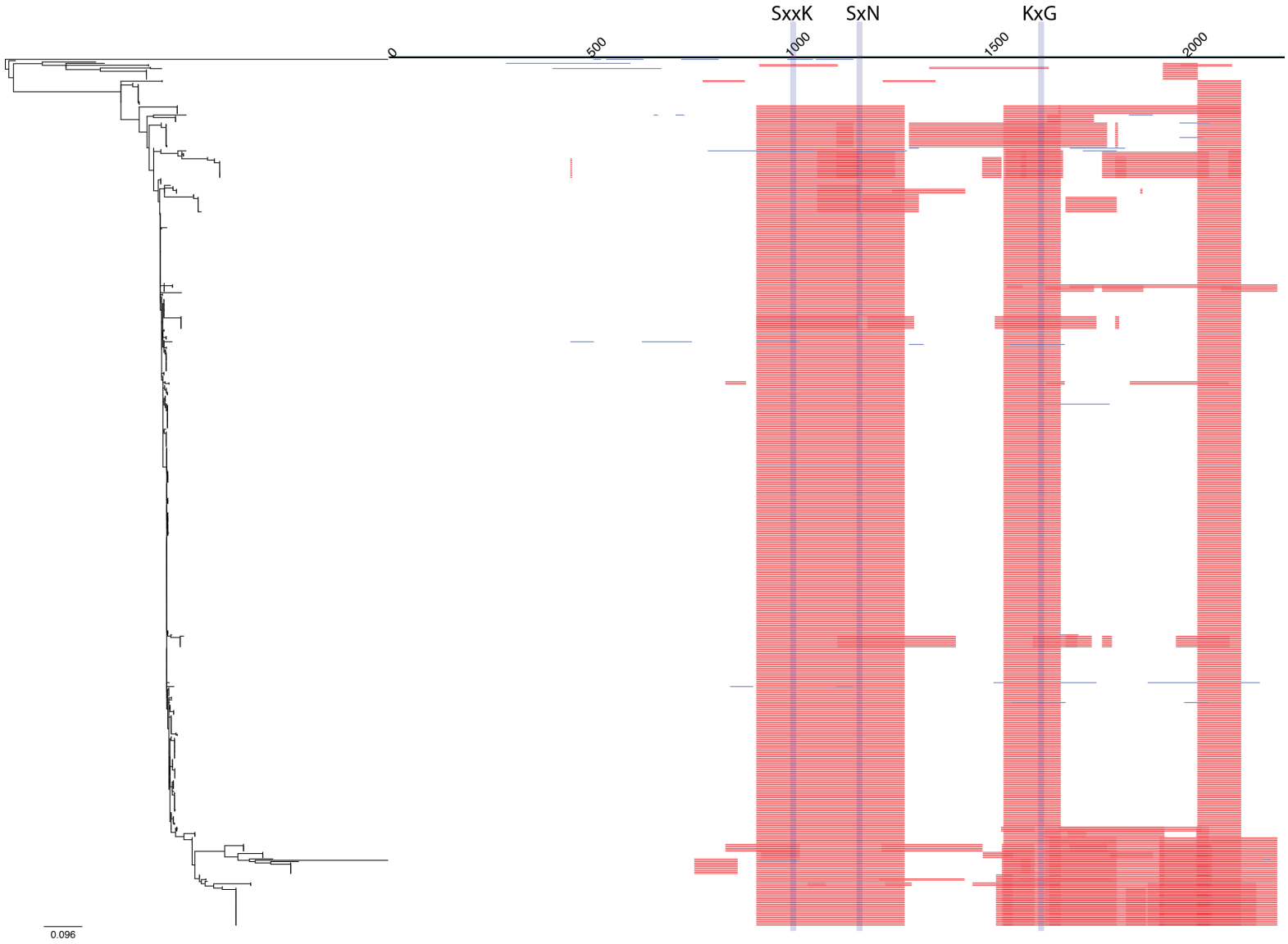


Figure 53: Recombination events predicted to have taken place in the Malawian *pbp2x* gene. These events are coloured red when common to two or more isolates, or blue when they are unique to an isolate. The positions of the key motifs (SxxK, SxN, and KxG) are indicated relative to their positions in this gene. A phylogeny, constructed in the absence of these events is also shown.

Recombination events were found to have affected all isolates within the sample set, with three distinct recombination events clearly identifiable as being highly conserved across the dataset (Figure 53). One of these events occurred within the region of the three SxN, and SxxK motifs, between nucleotides 889 and 1191 in the R6 *pbp2x* allele. The KxG motif appeared to have been affected by a separate recombination event, which was similarly highly conserved across study isolates. A number of additional recombination events were similarly found to overlap this conserved events. This indicated the occurrence of multiple recombination events that had affected the same locations (Stoltzfus et al., 1988), so that several different SNP densities had been detected at overlapping locations by the analysis tool.

5.3.2 *Pbp2b*

Substantially fewer recombination events (30) were identified in *pbp2b* compared to *pbp2x* (Figure 54). Removing recombination events from the alignment similarly lead to an improvement in the association between root to tip distances and isolation dates (pre $R^2=2.5x10^{-2}$ $p=3.67x10^{-2}$, post $R^2=2.53x10^{-2}$). Fewer SNPs were identified in *pbp2b* at 1734, of which 978 were found to have been introduced through recombination, giving an R/M ratio of 1.29 (2dp). This value is once again lower than might be expected as discussed previously (see 5.7.1).

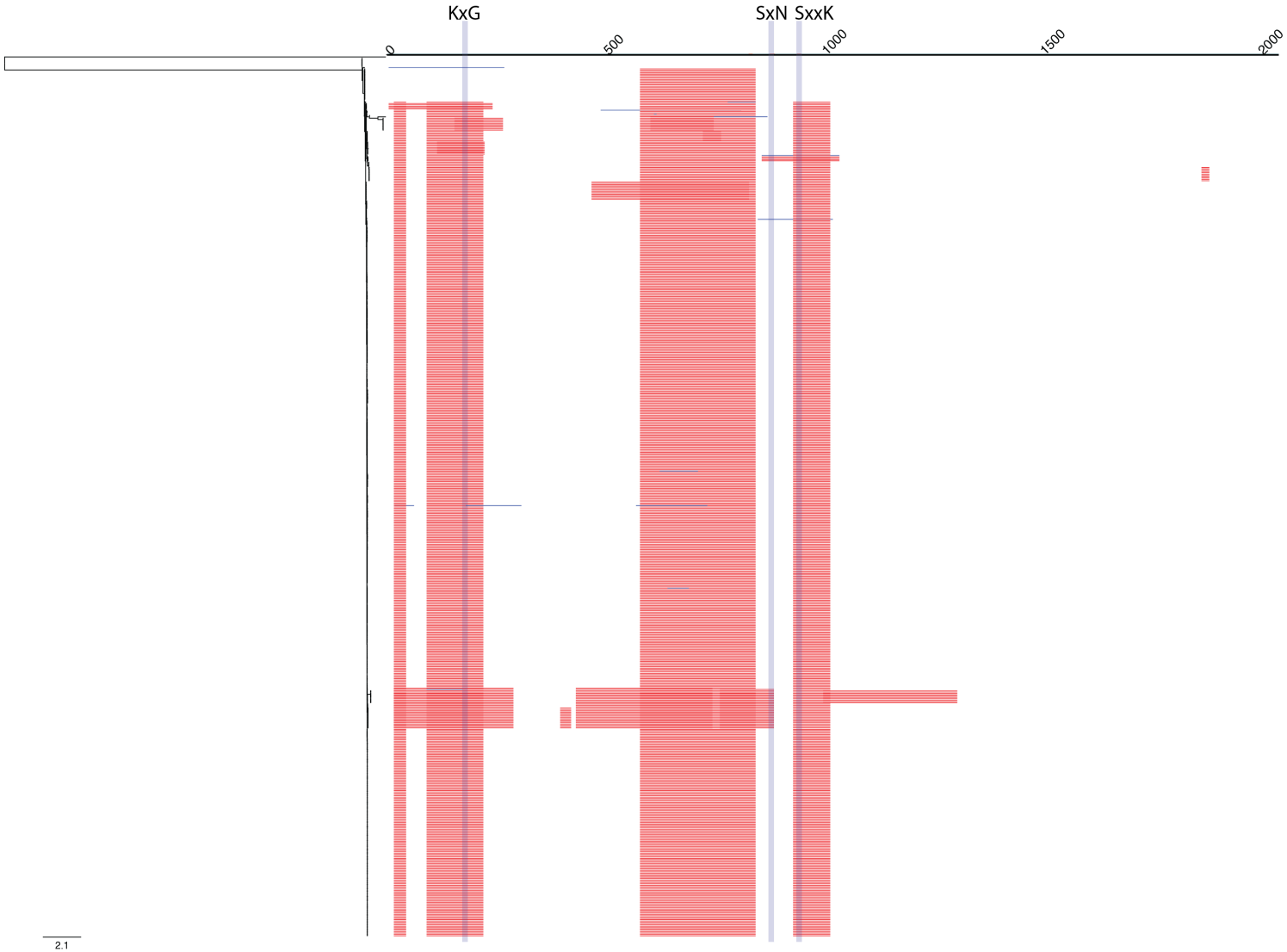


Figure 54: Recombination events predicted to have taken place in the Malawian *pbp2b* gene. These events are coloured red when common to two or more isolates, or blue when they are unique to an isolate. The positions of the key motifs (SxxK, SxN, and KxG) are indicated relative to their positions in this gene. Note, these motifs appear in reverse order as this gene occurs on the reverse DNA strand. A phylogeny, constructed in the absence of these events is also shown.

Similar to *pbp2x*, several widely conserved recombination events were found to occur in the *pbp2b* gene. Interestingly, whilst both KxG and SxxK motifs were affected by these events, very few recombination events were found to have affected the SxN location.

5.3.3 *Pbp1a*

A total of 1510 SNPs were identified within *pbp1a*, similar in number to those identified in *pbp2b*. Correspondingly only 49 recombination events were identified. Four of these events appeared to be widely conserved (Figure 55). Once again, one of these conserved blocks appeared to overlay the key mosaic regions of this gene. The R/M ratio was higher within this group at 4.18 with 1219 SNPs introduced through recombination events. In the case of *pbp1a*, however, removing the identified recombination events actually reduced the correlation between isolation date and root to tip distances (pre $R^2= 1.87x10^{-2}$, $p= 2.4x10^{-2}$, post $R^2= 9.24x10^{-3}$, p= 0.11). This suggests that recombination events may have been over represented within this dataset, with too much nucleotide polymorphism having been assigned to recombination. The branching pattern of the tree similarly suggested a lack of diversity within the main cluster of isolates. However, given that the mutation rate for pneumococci is estimated to be $3.3x10^{-6}$ substitutions per year, and the limited length of the gene (2160bp), this was not thought to affect the subsequent analysis of these events.

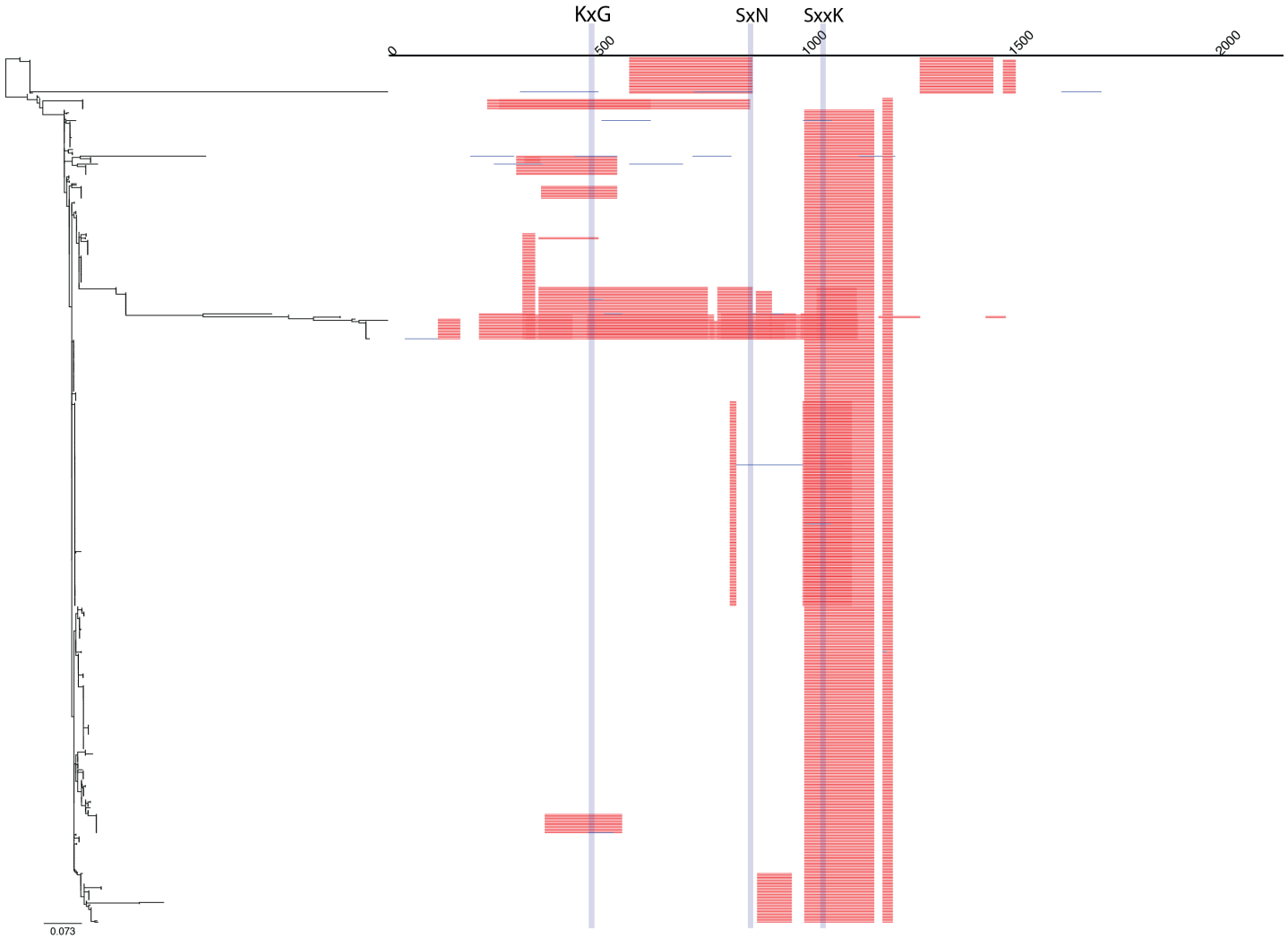


Figure 55: Recombination events predicted to have taken place in the Malawian *pbp1a* gene. These events are coloured red when common to two or more isolates, or blue when they are unique to an isolate. The positions of the key motifs (SxxK, SxN, and KxG) are indicated relative to their positions in this gene. Note, these motifs appear in reverse order as this gene occurs on the reverse DNA strand. A phylogeny, constructed in the absence of these events is also shown.

5.3.4 Additional Genes of Interest

Although not discussed previously *pbp1b* and *pbp2a* can carry mutations associated with changes in beta-lactam susceptibility *in vitro* (Denapaite et al., 2007, Job et al., 2008). Consequently these *pbp* genes were included in the recombinational analysis. Interestingly, these *pbp* alleles were far more conserved, with only 152 SNPs found in *pbp1b* and 338 SNPs identified in *pbp2a*. Only a single recombination event was subsequently identified to have occurred in *pbp1b*. Consequently, supporting the limited involvement of these genes in clinical resistance, recombination appeared to have played an insignificant role in the divergence of these genes.

DdlA is a D-alanine D-alanine ligase, involved in the attachment of these groups, which are subsequently attached to the UDP-MurNAc-pentapeptide during the cell wall synthesis pathway. Consequently, polymorphism within this gene could affect the attachment and construction of these branch chains, which *pbps* mimic. However, variation was once again limited within this locus with 603 SNPs identified across the sample set. Here only 3 recombination events were identified. Once again, this suggests a limited role of recombination within the divergence of this gene and promoter region.

MurF is responsible for attachment of D-alanyl D-alanine side chains to the UDP-MurNAc-pentapeptide, these side chains forming the substrate for PBP transpeptidation. Similar to DdlA, MurF could affect beta-lactam susceptibilities as a consequence of its involvement in attaching D-alanyl D-alanine side chains. The *murF* gene was found to be quite divergent, with 1266 SNPs identified, suggesting variation is tolerated within this allele. Correspondingly 34 recombination events were identified, however these were responsible for the introduction of only 444 SNPs. This suggests a large degree of the variation within this gene was not explained by the recombination events identified. This indicates recombination events effecting the whole gene may be important for increasing the variation of this gene as observed in this dataset- which was higher than would be expected based on estimates of mutations rates in this

species (Croucher et al., 2011), and was inconsistent with the other genes analysed.

MurM is involved in the serylation or alanylation of lysine in the pentapeptide side chains during cell wall synthesis, and has previously been found to affect beta-lactam resistance (Lloyd et al., 2008). However, no recombination events were identified within this gene, which was similarly highly conserved. This suggests MurM had little to no effect on beta-lactam susceptibilities within the dataset.

LeuS has previously been identified as a possible virulence factor (Hava and Camilli, 2002), but has not previously been identified as responsible for beta-lactam resistance clinically. Interestingly *leuS* carried 2293 SNPs, similar to the levels identified in *pbp2x*. In total 67 recombination events were identified, introducing 1036 of these SNPs (Table 28). Given the larger number of SNPs not identified as occurring due to recombination, once again, larger recombination events across the whole gene are likely to have occurred.

Gene	**Average**	**Median size**	**Max length**	**Min length**	**Count**	**Length of gene**
ddlA (with promoter)	61	61	64	50	3	1221 (with promoter)
leuS	134.73 (2dp)	127	580	7	67	2502
murF	81.53 (2dp)	63	233	18	34	1374
murM	0	0	0	0	0	1221
pbp1a	114.91 (2dp)	102	635	10	49	2160
pbp2a	0	0	0	0	0	2196
pbp2b	150.4	28.5	361	6	30	2043
pbp2x	131.89 (2dp)	107	499	4	103	2253
pbp1b	96	96	96	96	1	2466

Table 28: Summary statistics for recombination events identified within each gene. For each gene, the average length, median size, maximum and minimum lengths, total number and size of the gene, for perspective, are indicated.

5.3.5 Summary of Recombination Detection

As expected, recombination was found to have affected all three of the main *pbps* thought important for beta-lactam resistance. Interestingly, *pbp2x* was the most divergent in terms of SNPs: 3600 compared to ~1500 for *pbp2b* and *pbp1a*. A lack of recombination, and nucleotide polymorphism among *pbp's 1b*, and *2a* was consistent with a lack of involvement in clinical resistance for these genes.

Variation in the additional genes considered was also highly variable in terms of nucleotide polymorphisms present. Given the close locations of *murF* and *ddlA* to the highly recombinogenic *pbp2b*, it might be expected that recombination affects these loci at a similar rate, consistent with a hitchhiking effect (Enright and Spratt, 1999). These genes were found to be affected by recombination however to a lesser extent. This result would still be consistent with *pbp2b* being the focus of recombination, with variably sized recombination events occasionally introducing variation into these flanking genes. In addition, MurM modification appears to be highly limited. Based on the assumption of beta-lactam resistance being driven by the occurrence of recombination rather than point mutation, it therefore appears that MurM contributed insignificantly to beta-lactam resistance within the dataset.

5.4 Identifying the Source of Recombinations

The recombined blocks identified in this analysis were then compared to global collections for *S. mitis*, *S. oralis* and *S. pneumoniae* in order to determine the extent to which recombination occurs between these species and on a global scale.

5.4.1 Identifying gene mosaicism in Malawian isolates

In order to identify potential genetic donors, sequences from the identified recombination events were compared to publicly available datasets for the three species of interest: *S. mitis*, *S. oralis* and *S. pneumoniae*. For each species a database was compiled based on the sequences available from the European Nucleotide Archive (ENA, 2015). Each database contained a mix of reads from

whole genome studies, to studies that focussed on particular genetic loci. Due to wider interest in pneumococci generally, this database was the largest dataset, comprising of 13,578 reads. In contrast, far fewer sequences were available for *S. mitis*, 5380, and *S. oralis*, 2682. Consequently there was a strong bias in detecting shared sequences with other *S. pneumoniae* isolates.

The program blastall (section 2.7.3) was used to compare the recombined sequences that had been identified among the Malawian pneumococci, to the reads collated from the public archives. This approach was intended to try and identify the origin of the recombined fragments, i.e. whether these were unique to Malawi, or had originated from other globally circulating sources.

Each potential match was scored by blastall according to the quality (base-mismatches, gaps) and length over which the sequences matched (Figure 56). Blastall aligns each input read (query) to its internal sequence database (subject reads). During this alignment process gaps may be introduced into the sequence being aligned, or if only part of the query sequence aligns to the subject read, only the aligned section will be presented. The quality of the match (presence of nucleotide polymorphisms, introduction of gaps) between the aligned query and subject read is then recorded in the blastall output. Consequently, a query sequence may score highly in terms of quality, if it matched the subject well, but if only part of the query sequence has aligned, this will only be evident when considering the length of the alignment.

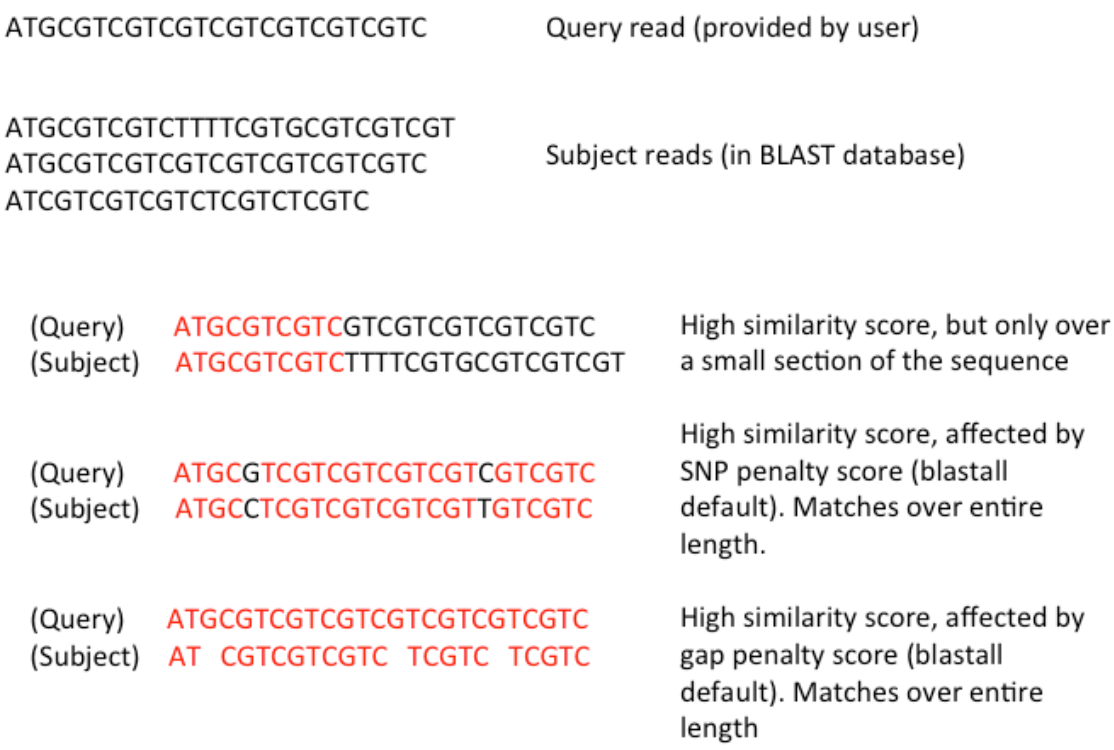


Figure 56: How an alignment is generated and scored in Blastall. A query is provided by the user, which Blastall attempts to align to sequences contained within it's internal database. Alignments are then scored according to the length, mismatches, and gaps occurring between the query, and subject reads being aligned. Consequently, whilst an alignment may score highly under one criteria, it may have a low score relative to another.

As such, Blastall indicates many possible alignments. To increase the certainty that the donor of the recombined sequence was identified, it was therefore necessary to add additional cut-off criteria. The cut-off criteria developed here was based on including Blastall matches of ≥90% in length and ≥90% in similarity.

Table 29 indicates for each gene, how the recombination events identified in the Malawian sequence datasets could be assigned to different species according to these cut-off criteria.

Gene	*S. mitis*	*S. oralis*	*S. pneumoniae*	Total in gene
ddla	1	1	0	2
leuS	12	2	44	58
murF	1	1	32	34
pbp1a	5	3	40	48
pbp1b	0	0	1	1
pbp2b	2	0	27	29
pbp2x	10	2	93	105
Total for species	31	9	237	277

Table 29: For each gene recombination events were assigned to donor species based on the Blastall criteria (sequences match ≥90 length and similarity). The results of the Blastall analysis are indicated, showing the number of recombination events that were assigned to each of the donor species. Also shown are the total counts for the number of events assigned to each species, and total number of events occurring in each gene tested.

In several cases the recombined sequences present within the Malawian pneumococcal dataset were found to share highest similarity with sequences present within the *S. mitis* and *S. oralis* datasets. These results therefore appear to support the circulation of all three species globally, owing to the high similarity identified between recombined sequences in the Malawian dataset, and the global collections represented by the blastall subject database.

The high similarity between recombined fragments within *pbp* genes in Malawi and the globally collated datasets, suggested that amino acid changes conferring reduced susceptibilities could have arrived within the Malawian pneumococcal population through international dissemination of resistance genes. To better determine the contribution of recombination to resistance in Malawi the recombination events shown in Table 29 were compared to determine whether these events had introduced SNPs associated with changes in beta-lactam susceptibility based on the literature. In addition, a putative genetic donor was identified based on the highest score in terms of similarity and length. The results from this analysis are presented in Table 30.

For *pbp2x*, *pbp2b* and *pbp1a* genes amino acid changes associated with reductions in beta-lactam susceptibilities were identified within the recombination events identified in Table 29. Interestingly, although recombination events were more frequent among *pbp2x* alleles, such events appeared to rarely introduce clusters of SNPs associated with a loss in beta-

lactam susceptibility. In contrast, up to seven resistance-associated SNPs were introduced in a single recombination event for *pbp1a* according to this study. Furthermore, there was little indication that these recombinations had been derived from *S. mitis* or *S. oralis* species. This could indicate a limited role for recombination with these two species in the development of beta-lactam resistance. Instead, whilst such resistance may originate in *S. mitis* and *S. oralis*, further modification of these resistance sequence blocks may occur within the pneumococcal population. Greater circulation and contact of pneumococci could then facilitate the spread of such resistance material once acquired under this hypothesis. However, the much greater availability of pneumococcal sequence for analysis compared to *S. mitis* and *S. oralis* could similarly lead to more recombination events associating with pneumococci due to the dataset bias highlighted previously.

Gene	Recombination event	Amino acid introduced	Originator
pbp2x	923_1179	R384G, T338A	*S. pneumoniae*
pbp2x	923_1296	S389L, T338A, R384G	*S. pneumoniae*
pbp2x	1076_1181	R384G	*S. pneumoniae*
pbp2x	1076_1185	R384G	*S. pneumoniae*
pbp2x	1124_1167	R384G	*S. pneumoniae*
pbp2x	1124_1272	R384G	*S. pneumoniae*
pbp2x	1127_1425	R384G	*S. pneumoniae*
pbp2x	1491_1540	N514H	*S. pneumoniae*
pbp2x	1493_1568	N514H	*S. mitis*
pbp2x	1520_1779	Q552E, D567N, N514H	*S. pneumoniae*
pbp2x	1523_1779	Q552E, D567N, N514H	*S. pneumoniae*
pbp2x	1526_1568	N514H	*S. pneumoniae*
pbp2x	1539_1623	N514H	*S. pneumoniae*
pbp2x	1545_1689	V516I, Q552E	*S. pneumoniae*
pbp2x	1619_1767	Q552E, D567N	*S. pneumoniae*
pbp2x	1649_1773	Q552E, D567N	*S. pneumoniae*
pbp2x	1649_1812	Q552E, D567N	*S. pneumoniae*
pbp2x	1649_1950	Q552E, D567N	*S. pneumoniae*
pbp2x	1652_1699	Q552E, D567N	*S. pneumoniae*
pbp2x	1712_2121	N605T	*S. pneumoniae*
pbp2x	1793_1897	N605T	*S. pneumoniae*
pbp2b	432_745	E476G, T446A	*S. pneumoniae*
pbp2b	672_871	T426K, Q427L	*S. pneumoniae*
pbp2b	579_844	E476G, T446A, T426K, Q427L	*S. pneumoniae*
pbp1a	236_871	S462A, F577Y, Q576G, S575T, A585V, T574N, L583M	*S. pneumoniae*
pbp1a	265_631	F577Y, Q576G, S575T, A585V, T574N, L583M	*S. pneumoniae*
pbp1a	306_550	F577Y, Q576G, S575T, A585V, T574N, L583M	*S. pneumoniae*
pbp1a	315_505	F577Y, Q576G, S575T, A585V, T574N, L583M	*S. pneumoniae*
pbp1a	360_505	F577Y, Q576G, S575T, A585V, T574N, L583M	*S. pneumoniae*
pbp1a	360_769	F577Y, Q576G, S575T, A585V, T574N, L583M	*S. pneumoniae*
pbp1a	366_550	F577Y, Q576G, S575T, A585V, T574N, L583M	*S. pneumoniae*
pbp1a	375_562	F577Y, Q576G, S575T, A585V, T574N, L583M	*S. pneumoniae*
pbp1a	579_877	S462A, P432T, E397I, I459M	*S. mitis*
pbp1a	732_826	S462A, I459M	*S. pneumoniae*
pbp1a	735_877	S462A, P432T	*S. pneumoniae*
pbp1a	792_877	P432T	*S. pneumoniae*
pbp1a	801_982	P432T	*S. pneumoniae*
pbp1a	1002_1171	T371A	*S. pneumoniae*
pbp1a	1032_1129	T371A	*S. pneumoniae*

Table 30: Recombination events that potentially imported beta-lactam resistance associated amino acid changes. Also indicated is the species from which the recombination was thought to have been derived, based on the previous Blastall analysis.

5.4.2 Summary of identifying sources of recombination events

This section describes an attempt to identify the potential donors of the recombination events from a collection of Malawian pneumococcal isolates. By using the sequence databases available for *S. mitis*, *S. oralis* and *S. pneumoniae* collected globally, outside of Malawi, the role globally circulating populations of these three species has had on the development of beta-lactam resistance in Malawi was investigated. The ability to identify cases where recombined sequences shared a high degree of homology to those collected globally acts in support of this hypothesis. Furthermore, it is clear that recombination with these globally circulating populations has the potential to introduce beta-lactam resistance into the Malawian population. Given the limited antibiotic usage in Malawi, it is possible that beta-lactam resistance first arose among the globally circulating populations. Recombination between the Malawian pneumococcal population and these resistant isolates subsequently could have allowed the transfer of alleles carrying amino acid alterations affecting beta-lactam affinities to enter this population. Depending on the local selective pressures, these alterations would either be subsequently selected for or purged from the population.

5.5 Discussion

In this chapter the role of recombination in beta-lactam resistance in Malawi, and the potential role international recombinational exchanges have had on the emergence of beta-lactam resistance in Malawi has been described. *Pbp* divergence appears to be an important indicator for beta-lactam resistance, however, it has been demonstrated that the location of the recombined fragment is also important for the overall affect. The recombined fragments identified frequently extended beyond the normal SxN, K(T)G and SxxK motifs, or occurred in other regions of the *pbp* genes. This suggests that little of this variation affects the beta-lactam MIC, which is supported by the finding that branch length, (root-to-tip distance) correlates poorly with the MIC recorded for each strain. Some clustering between strains of similar MICs was identified, indicating shared divergent material. However, penicillin resistance was not restricted to

particular isolates, suggesting that a number of different beta-lactam resistance mechanisms are currently contributing to losses in beta-lactam susceptibility in Malawi.

For *pbp1a*, variation as measured by root to tip distance, plateaued at a much lower level when compared to the other *pbp* alleles. This suggests that *pbp1a* was more conserved relative to *pbps 2x* and *2b*. This plateauing largely resulted from two divergent clusters of penicillin resistant isolates that shared genetic variation at this locus. In one case, these were found to belong to ST63 and commonly ST14, whereas the second cluster contained isolates from a variety of STs, although commonly carried a serotype 19F. The lack of further divergence within *pbp1a* could indicate that this gene is less tolerant to divergence than *pbp2x* and *2b*. However, the degree to which *pbps* can vary without losses in fitness is currently untested. Furthermore, compensatory mutations at different genomic loci could facilitate tolerance to *pbp* variation (Orio et al., 2011). Consequently, the presence of other mutations genome-wide could similarly affect the degree of *pbp* variation tolerated.

5.5.1 Identifying recombinations

Recombination was identified in the *pbp2x* gene far more frequently than was identified in the *pbp1a* gene, consistent with pp2x variation being necessary for the initial loss of beta-lactam susceptibility, with *pbp1a* changes necessary for high-level beta-lactam resistance (Smith and Klugman, 1998). Interestingly however, recombination was identified to have occurred in the *pbp1a* gene more frequently than in the *pbp2b* gene. This suggests that a greater selective pressure exists on the *pbp1a* gene. This is consistent with the antibiotic usage in Malawi. Whilst penicillin usage should drive recombination across *pbp1a*, *2b* and *2x*, (Smith and Klugman, 1998), ceftriaxone resistance should drive recombination in *pbp1a* and *2x* alone (Munoz et al., 1992). This is consistent with ceftriaxone usage having been sufficient to drive recombination among Malawian pneumococci, also demonstrated by the loss in ceftriaxone susceptibility that has been observed recently (pers. comm. Everett 2015).

5.5.2 Speculative resistance genes

Recombination events were similarly identified within genes speculated to be involved in beta-lactam resistance. *DdlA* and *murF* were found to have undergone recombination, although given the proximity of these genes to *pbp2b* it is likely that recombination can span these genes (Enright and Spratt, 1999, Croucher et al., 2012). Interesting MurM, similar to PMEN1 was highly conserved, although this gene has previously been found to confer beta-lactam resistance (Smith and Klugman, 2001, Lloyd et al., 2008). The similarity of MurM to the susceptible R6 strain indicates MurM modification has not contributed to the beta-lactam resistances in Malawi at present. *LeuS* similar to PMEN1 was found to have been subject to frequent recombination. Interestingly some of these recombination events showed high sequence similarities to *S. oralis* and *S. mitis* further supporting the possible involvement of this gene in beta-lactam resistance. The role of this gene in beta-lactam resistance however is currently only speculated.

5.5.3 Origins of recombinations

Identifying sources of recombination is complicated by sequence divergence, the subsequent spread of resistance determinants following initial acquisition, and determining the direction of transfer. Here recombined fragments were compared to publicly available databases at comparative loci in order to identify putative recombinational donors. In order to filter the number of results a strict cut-off value was also applied, ensuring that sequence matches were highly similar in terms of length and similarity. Consequently, only more recent or highly conserved recombinations are likely to have been identified. Whilst this limits the extent of this analysis, this cut-off ensures greater confidence that the matches identified did not occur by chance alone. “Origin” in this case was then inferred, based on identifying divergent sequence blocks present at similar loci in different strains. The genetic donors in this study appeared to far more frequently involve other pneumococci, than *S. mitis* or *S. oralis.* This was perhaps due to the far greater number of pneumococci available for analysis compared to

these other mitis group species. Consequently, whilst this result could indicate that resistance in Malawi has been driven to a far greater extent by intra- rather than interspecies recombination, the sample biases could equally explain this finding.

An extension of this analysis could have assessed the degree to which fragments, present in the donor, differed from the background SNP frequency- i.e. were these fragments also divergent in the identified donor sequence? Whilst this was not found to be the case visually, the high level of divergence among *S. oralis* and *S. mitis,* and the limited number of isolates available for study prevented a more rigorous comparison from being made.

5.5.4 Recombination and beta-lactam resistance

Whilst a number of possible recombination "donors" were identified, few recombined fragments in these cases were found to have lead to the acquisition of resistance associated SNPs in the Malawian pneumococci. Although *pbp2x* was the most divergent gene, few of the recombined blocks appeared to have introduced more than one resistance associated SNP. In some cases (e.g. R384G) the same mutation appeared to have been introduced multiple times. As such, this could suggest that some mutations may be more widespread routes to beta-lactam resistance than others.

Very little of the variation introduced into *pbp2b* was found to have been associated with the introduction of resistance associated SNPs, and again, no common resistance pathway was identified, with multiple different resistance associated SNPs having been introduced in different strains. Interestingly, although *pbp1a* had been identified as the least divergent *pbp* gene, a high majority of the SNPs present within recombined fragments were identified to contain resistance-associated mutations. Mutations that occurred around the KTG motif appeared particularly important for resistance to occur in this gene. KTG modification can affect the ability of beta-lactams to access the *pbp's* active

site (Sauvage et al., 2008), suggesting that this could be an important mechanism for resistance among Malawian *pbp's*.

5.6 Conclusions

The analysis presented here indicates a role for global transmission of resistance blocks of sequence in the development of beta-lactam resistance in Malawi. No one mode of resistance appears to be common, but rather multiple independent recombination events appear to have occurred, having differing affects on beta-lactam resistance dependent on where in the *pbp* alleles they occur, and the number of resistance associated SNPs being introduced. Despite the high numbers of recombination events that have affected *pbp2x* and *pbp2b*, little of the variation introduced appears to encode for beta-lactam resistance, either occurring in locations thought less important for enzyme-substrate interactions, or containing SNPs that have not been found to affect beta-lactam resistance *in vitro*. Recombination with members of the mitis group appears to occur widely, with other genes, hypothesised to affect beta-lactam susceptibility similarly recombining with these species. The degree to which such recombination occurs genome-wide however could not be conducted in part due to a lack of whole genome sequences for *S. mitis* and *S. oralis*. The high level of sequence diversity present within the mitis group also poses problems for phylogeny and recombination detection algorithms that rely on comparison of a consensus genome sequence between isolates.

Recombination with mitis species circulating globally appears to have contributed to beta-lactam resistance in Malawi. Recombination appears to occur far more frequently however with other, globally circulating pneumococci, rather than *S. mitis* or *S. oralis.* The current lack of sequence data for these latter two species though, means that this result could be an artefact of sample size. If on the other hand this result is true, this could suggest beta-lactam resistance has been acquired rarely by pneumococci, and that subsequent modification and frequent exchanges between pneumococci has facilitated its spread.

The increasing reliance on ceftriaxone would be expected to reduce the pressure on *pbp2b* modification and increase *pbp1a* and *pbp2x* modification. Interestingly, although high penicillin resistance is rarely identified, *pbp1a* was rich in resistance associated SNPs, and branching patterns were also different from *pbps 2b* and *2x*. Consistent with an enhanced selective pressure on this allele, recombination was similarly found to have affected this locus more frequently than that of *pbp2b*.

This analysis continues to indicate a high level of strain mixing with Malawi, with no single mode of resistance predominating as yet. In only *pbp1a* was there any indication of conserved resistance mechanisms. Here two groups of divergent isolates were apparent. Furthermore, whilst these groups appeared to have two different origins, several clusters of SNPs were shared between isolates from these separate groups. Such lineages, particularly the ST63 group could therefore be important in the future emergence of beta-lactam resistance in Malawi.

6 Oxidative Stress

Free oxygen can be problematic for organisms as it can yield reactive oxygen intermediates (ROIs) capable of unrestricted oxidation. When left uninhibited such molecules are capable of causing rapid cell death (Figure 57)(Stohs and Bagchi, 1995, Finkel and Holbrook, 2000, Droge, 2002). ROIs can result from the excitation of O_2 leading to the formation of an oxygen singlet (O_2^1) or the transfer of electrons to O_2, resulting in a superoxide radical (O_2^-), hydrogen peroxide (H_2O_2), or a hydroxyl radical (HO-)(depending on whether one, two or three electrons are transferred respectively). Hydrogen peroxide is particularly potent as it is capable of reacting with the free iron within the cell. When this free iron occurs close to the cellular DNA, hydroxyl radicals can be formed which oxidise base and ribose moieties (Imlay, 2013). Such reactions have been hypothesised to result in tandem duplications, DNA shearing and increased mutation (Pericone et al., 2002).

ROI formation can originate from internal and external sources such as the inhibition of replication machinery or exposure to damaging UV light and antibiotics. ROIs can also be derived from other organisms, some organisms availing of its potency to attack co-colonisers (Pericone et al., 2000). Consequently, sources of oxidative stress are ubiquitous and all organisms must develop mechanisms to tolerate this form of attack. Oxygen scavenging molecules, the compartmentalisation of metabolic processes that yield ROIs, and the active protection of sensitive intracellular components are frequently employed mechanisms to guard against these damaging agents (Pericone et al., 2003, Cooke et al., 2003, Kohanski et al., 2007).

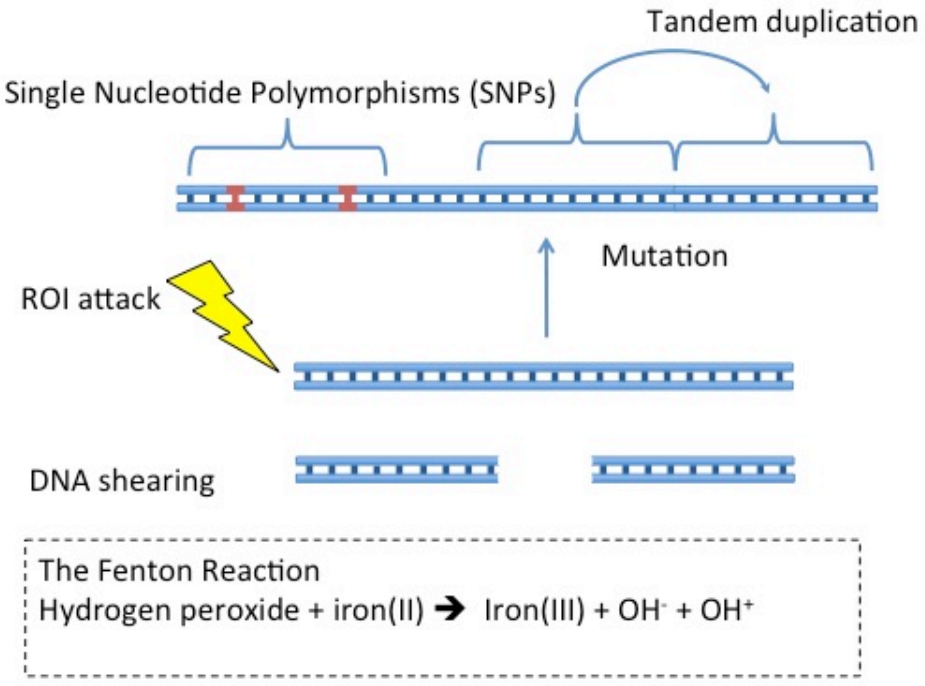


Figure 57: Types of DNA damage hypothesised to result from attack by reactive oxygen species. The occurrence of DNA shearing (double strand breaks) is potentially very harmful to the organism. In such cases it is essential that the damaged DNA is re-joined. Such events are also thought to result in the deletion of sections of DNA. The Fenton reaction, which can occur within the cell is thought an important source of DNA damaging free radicals.

Organisms, such as the oral streptococci, are constantly in contact with free oxygen and as such they are particularly exposed to ROIs. Utilising oxygen through aerobic respiration allows such organism to gain additional ATP through pyruvate metabolism (Carvalho et al., 2013). However, lacking a complete tricarboxylic acid cycle (TCA), they are unable carry out the complete reduction of glucose to oxaloacetate. Pyruvate is produced by conventional glycolysis and subject to the competing activities of lactate dehydrogenase, pyruvate formate-lyase, pyruvate oxidase and the hypothetical pyruvate dehydrogenase complex. Unable to complete the metabolism of pyruvate, the process is rendered far less efficient than conventional TCA metabolism, and leads to the production of potentially toxic hydrogen peroxide as a by-product (Holzapfel and Wood, 1995, Taniai et al., 2008, Carvalho et al., 2013). Although more stable than free oxygen, hydrogen peroxide is toxic to the cell in the presence of ferrous iron (Fe^{+2}), where the Fenton reaction can occur yielding hydroxyl radicals. Many aerobic bacteria possess the enzyme catalase, which acts to reduce hydrogen peroxide to water and oxygen outside of the cell. Lacking this

enzyme, hydrogen peroxide instead diffuses out of the cell and accumulates in the supernatant, where it has been found to reach concentrations in excess of 1mM (Pericone et al., 2000).

Hydrogen peroxide concentrations of between 0.1 and 1.0mM are shown to be toxic to many species of bacteria (Repine et al., 1981, Dowds and Hoch, 1991, Pericone et al., 2000), these concentrations showing a similar level of toxicity to that produced by neutrophilic oxidative bursts per gram of total cellular protein (Duane et al., 1993).

The oral streptococci are found to have a high tolerance for hydrogen peroxide, which has been hypothesised to aid biofilm formation, genetic exchange (Auzat et al., 1999, Kreth et al., 2009) and colonisation- by inhibiting the growth of other niche colonisers (Kreth et al., 2009). *S. pneumoniae* is found to consistently reduce the growth of *H. influenzae* when grown *in vitro*, which Pericone and colleagues (2000) attributed to hydrogen peroxide inhibition. Other nasopharyngeal colonisers such as *S. aureus, N. meningitidis* and *Moraxella catarrhalis* were similarly found to be susceptible to the lytic properties of hydrogen peroxide, despite their ability to produce catalase (Bisaillon et al., 1985, Singh et al., 1997, Pericone et al., 2000, Pericone et al., 2003, Selva et al., 2009, Margolis et al., 2010). This hypothesis is similarly supported by *in vivo* observations where despite *S. pneumoniae* and *H influenzae* being the most commonly isolated pathogens from the nasopharynx, (Del Beccaro et al., 1992, Klein, 1997), the frequency with which they co-occur is far lower than would be expected based on these prevalences (May, 1954, Luotonen, 1982, Pericone et al., 2000).

The pneumococcus does not possess many of the standard mechanisms for ameliorating oxidative stress, lacking the major hydrogen scavenger catalase (Hoskins et al., 2001, Tettelin et al., 2001), as well as the stress inducible OxyR or PerR regulons of many other species (Christman et al., 1989, Horsburgh et al., 2001). Other ROS scavengers, such as alkyl hydroperoxidase (Paterson et al., 2006), superoxide dismutase (SodA) (Yesilkaya et al., 2000), and NADH oxidase

(Auzat et al., 1999) are instead present. The pneumococcus also possesses widely conserved antioxidant thioredoxin enzymes that can repair readily oxidised cysteines and methionines (Ezraty et al., 2005, Collet and Messens, 2010) helping to protect internal protein function. The pneumococcus is not however impervious to hydrogen peroxide induced damage, and mutation rates are estimated to be between 10 and 200 fold higher in the presence of hydrogen peroxide, although growth was not found to be inhibited in this study (Pericone et al., 2002). Pericone and colleagues (2002) found endogenous hydrogen peroxide was responsible for the proliferation of a number of structural variants within the genome of stressed pneumococci, including tandem duplications. Interestingly, spontaneous tandem duplications within the pneumococci's capsule locus were found to lead to gene inactivation, and unencapsulation (Waite et al., 2001, Waite et al., 2003). Slipped-strand repair resulting in the removal of these deletions was found to occur at a log linear rate, depending on the length of the repeat (Figure 58)(Waite et al., 2001).

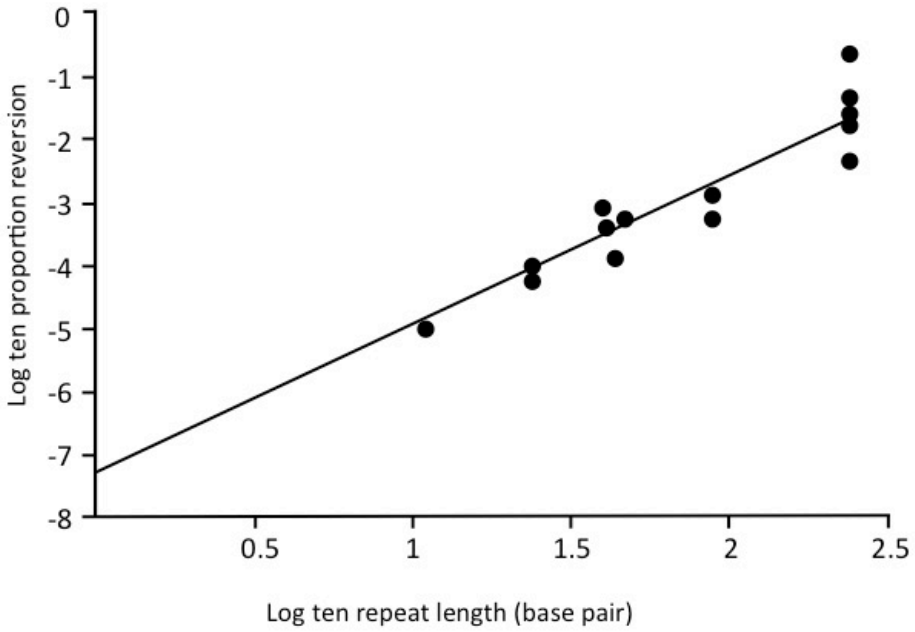


Figure 58: Phase variation in pneumococci. SCVs were found to revert on a log scale, proportional to the size of the repeat region, likely due to the effects of slip-strand repair in removing tandem repeats (from Waite et al., 2001).

Although the link between genomic damage and hydroxyl attachment is well established, the extent of damage and the mechanisms that allow the

pneumococcus to survive such levels of toxicity are not fully understood. Furthermore, whilst the Fenton reaction can be regulated by the availability of ferrous iron, this ion is found to occur in similar concentrations between the pneumococci and other bacterial species (Pericone et al., 2003). Instead, understanding the ability for pneumococci to withstand oxidative stress has focussed on the roles of NADH oxidase and the pyruvate oxidase, SpxB. NADH oxidase is found to be required for *in vitro* growth and acts to deplete available oxygen by four-electron reduction to H_2O (Auzat et al., 1999, Pericone et al., 2003). The SpxB enzyme, responsible for the production of hydrogen peroxide, also seems important for managing the toxic effects of this by-product (Pericone et al., 2003). Pneumococcal mutants possessing a non-functional SpxB gene have been found to suffer a 10^2 to 10^3 fold greater killing rate at 20mM exogenous hydrogen peroxide levels relative to the wild-type (Pericone et al., 2003).

In addition to managing the effects of hydrogen peroxide, little is known about the variation between strains in their ability to produce this toxin. A competitive advantage could be offered to pneumococcal lineages that can produce higher concentrations of hydrogen peroxide, inhibiting the growth of co-colonisers more effectively than in the case if lower concentrations of this toxin were produced. However, such lineages could suffer greater rates of hydrogen peroxide damage as a result. Consequently, the aim of this study was to assess the level of genome-wide damage caused by continuous exposure to hydrogen peroxide, and to identify the levels of hydrogen peroxide produced within a globally important pneumococcal lineage.

6.1 Project Aims

- To identify the genome-wide effects of continuous exposure to hydrogen peroxide
- To measure the variability in hydrogen peroxide production within PMEN1

6.2 Study Samples

Three strains were used for biofilm and batch culture analyses (Table 31). Strains 37A and 0100993 carried a serotype 3 capsule, whereas strain A42174 carried a serotype 1 capsule. All isolates were derived from human samples: strains 0100993 (Throup et al., 2000) and A42174 were isolated from blood, whilst strain 37A was isolated from the throat. Using the MLST typing scheme (pubmlst.org) (Enright and Spratt, 1998) 37A and 0100993 strains were assigned to the sequence type 180. This is one of the most prevalent serotype 3 sequence types worldwide. Between 1984 and 1998 it was responsible for multiple occurrences of pneumonia, meningitis and bacteraemia across Europe, Canada and Taiwan (Waite, 2001). More recently the introduction of the heptavalent conjugate vaccine has favoured a return of this non-vaccine type strain, which has had a resurgence in America (Beall et al., 2006, Reis et al., 2008), Europe, and Japan (Isozumi et al., 2008, Croucher et al., 2013).

Strain A42174 was assigned to the ST217 and was isolated from Malawi in 2006. Serotype 1 pneumococci are a frequent cause of invasive pneumococcal disease among children and adults and are often associated with high virulence (Ramdani-Bouguessa and Rahal, 2003, Martens et al., 2004, Adegbola et al., 2006), although in Malawi it is rarely drug resistant. This ST is also prevalent in pneumococcal disease in Malawi, the particular isolate used here sourced from the blood of a HIV positive male.

Strain	Origin	Serotype	Source	MLST
A42174	Blantyre, Malawi	1	Blood	217
0100993	London, UK	3	Blood	180
37A	Oxford, UK	3	Throat	180

Table 31: Pneumococcal isolates used in this study, indicating their origin, serotype, source and ST.

6.2.1 Capsule locus

The capsule is an important virulence factor, and down-regulation appears to be an important during nasopharyngeal colonisation. The capsule covers sensitive immunogens, which are uncovered in the unencapsulated state. Among these are adhesions, which must interact with receptors on the epithelial cells surface in

order for colonisation to take place. Two capsule types were included in this study. Strains 0100993 and 37A represent serotype 3 strains. This serotype was favoured by Waite (Waite, 2001) owing to the large mucoid serotype 3 capsule. Unencapsulation consequently leads to a clear change in colony morphology readily observable when colonies are cultured on a blood agar plate.

In contrast, strain A42174, a serotype 1 strain, does not produce an observable capsule. However, Waite and colleagues (2003) demonstrated how tandem duplication changes within the capsule locus commonly occurred in a non-serotype 3 isolates. As such, serotype 1 would be expected to undergo similar changes in tandem duplications within its capsule locus. Furthermore, serotype 1 is rarely observed in carriage, despite its prevalence in disease. As such, an ability to switch between an encapsulated and unencapsulated form could explain the lack of detection during normal colonisation. Consequently, the ability of serotype 1 strains to become unencapsulated was of interest, as well as its tolerance to hydrogen peroxide stress.

6.2.2 Hydrogen peroxide production

Hydrogen peroxide was produced at a similar rate by all strains when grown in batch culture (Figure 59). Hydrogen peroxide was similarly not detected when isolates were grown under anaerobic conditions, reflecting the change in mode of sugar metabolism to one independent of the SpxB enzyme. Figure 59 also demonstrates that hydrogen peroxide was rapidly generated up to a concentration of 1mM, the upper limit of detection after approximately 15 hours, by which time strains had reached stationary phase. Dilution series analysis, which was also carried out, indicated that all three strains had an MIC of 4mM, when grown in the presence of hydrogen peroxide.

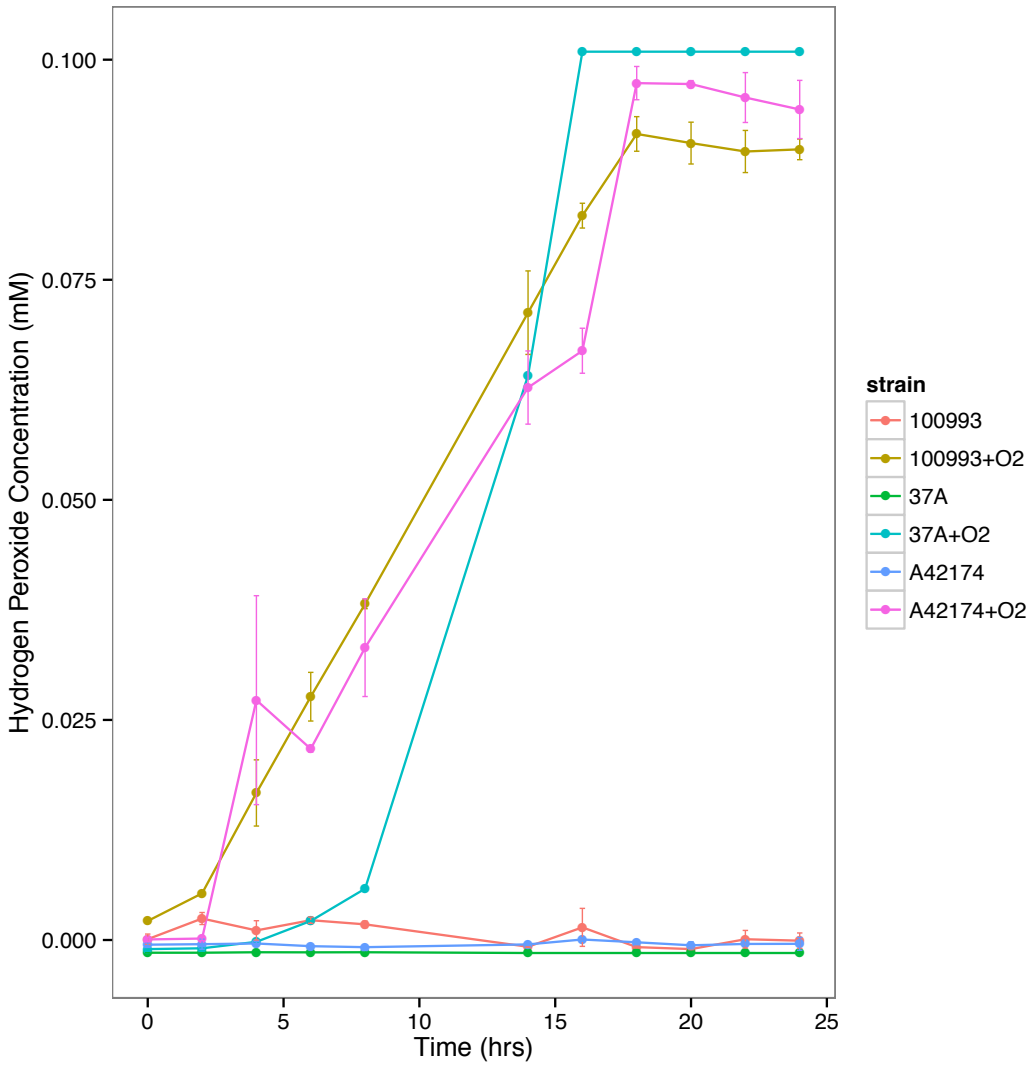


Figure 59: Hydrogen peroxide production under aerobic and anaerobic conditions for study isolates grown in batch culture. See methods for calculations.

6.2.3 Biofilm Growth

During planktonic growth in batch culture, pneumococci typically undergo autolysis once stationary phase has been reached, after approximately 16 hours of growth (Regev-Yochay et al., 2007). In contrast, nasopharyngeal carriage can occur for periods of 6 months (Gray et al., 1980). The sorbarod method was developed to study pneumococcal populations under conditions more closely resembling the nasopharyngeal environment. Populations can be established for periods over 27 days, with temperature, growth media, and atmosphere adjustable in the system (Hodgson et al., 1995, Budhani and Struthers, 1998, Waite, 2001). The sorbarod apparatus were set up as described previously (see

methods), with BHI media delivered to the sorbarod filter at a rate of 0.1ml/min. Experiments were carried out under aerobic conditions, at 34°C to better model the nasopharyngeal temperature. Sorbarod filters were inoculated with bacteria at $1x10^6$ CFU and incubated 48 hours, sufficient for a biofilm to be formed (Waite, 2001). Samples were then taken on conclusion of the experiment by collecting media effluent dripped from the sorbarod filter. Hydrogen peroxide was delivered in the growth media to enhance the effects of oxidative stress. Experimentation with 0.25mM hydrogen peroxide did not result in SCV production after 48 hours of growth, consequently the exogenous hydrogen peroxide concentration was increased to 2mM found sufficient for SCVs to be identified and isolated over this period (Figure 60).

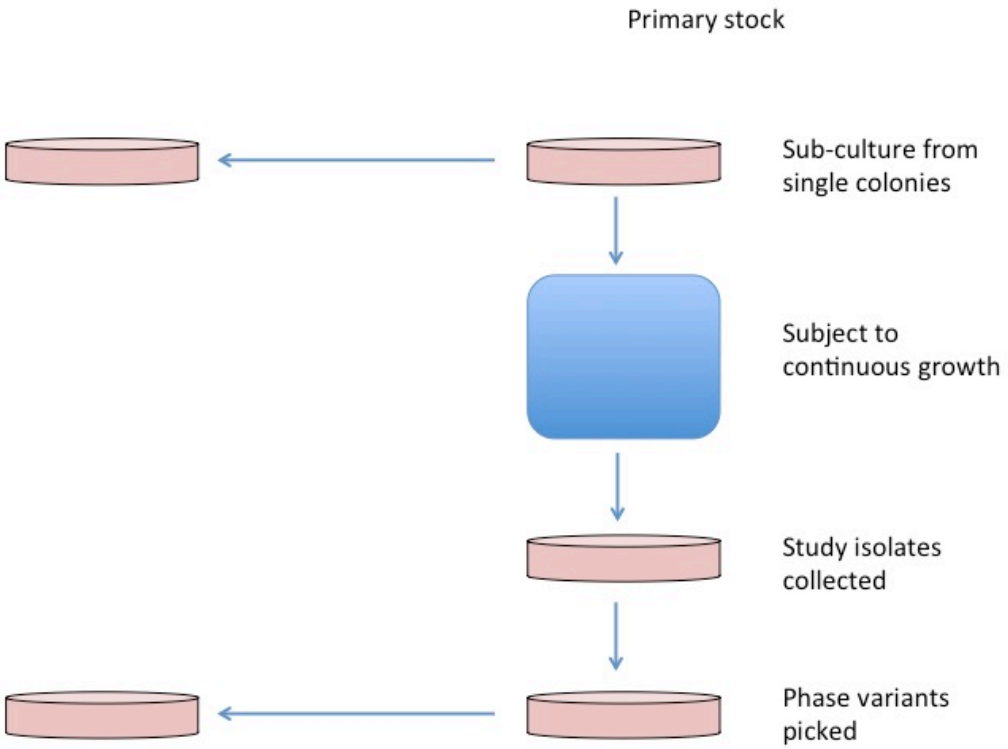


Figure 60: Study design. Two control isolates were sampled and sequenced prior to growth in the sorbarod apparatus. SCV isolates were identified following culture of the effluent, collected after 48 hours of growth in the sorbarod. Two SCVs were chosen for sequencing.

In addition, to limit the effects of hydrogen peroxide production in the sample populations, when isolates were grown on agar plates, these were incubated

anaerobically, using candle-jars containing Campygen sachets, to remove exogenous oxygen, at 37^{o}C.

6.2.4 Phase Variants

Phase variants were observed in both serotype 3 populations studied. Phase variation occurred least in strain 0100993, with SCV's accounting for approximately 6% (171 of 2974) of the plated colonies. The 37A population carried a higher occurrence of SCV's, which occurred at a rate of approximately 11% (194 of 1786).

SCV's were sampled and reversion rates calculated. For strain 0100993 it was possible to isolate SCV's, which did not revert to the encapsulated form whether grown under aerobic or anaerobic conditions. In contrast strains 37A reverted at a high frequency. The SCV colonies isolated from strain 0100993 were found to revert at a rate of approximately 30%. For strain 37A, SCV's reverted at a rate of over 60%, and it swas not possible to maintain populations of the unencapsulated forms over multiple generations.

Phase variation was not investigated for A42174 isolates, as the capsule is unobservable when grown on a plate. Instead, abnormally small isolates were chosen for sequencing.

6.2.5 Extraction and Sequencing

For the sequencing analysis two single colonies were isolated from the starting population, which are referred to subsequently as control isolates. Two SCV's were sampled from the media effluent following incubation in the sorbarod apparatus. SCV's are representative of bacteria growing on the sorbarod, and furthermore indicated isolates that had undergone mutation. Consequently in order to assess the genome-wide effects of this damage, these isolates were deliberately chosen for sequencing. In addition an attempt was made to isolate non-reversible SCV's, as tandem duplications, likely responsible for reversible phase variation, are known to cause difficulties for genome assembly. Whilst this

was achieved for 0100993 strains, it was not possible to isolate such non-reversible acapsular isolates for strain 37A, due to the high frequency of phase reversion.

Phase reversal following culture was particularly problematic as in order to grow sufficient quantities of bacteria for sequencing, multiple plates were used, and in some cases batch culture in 15ml tubes was used to supplement this. Plate growth was used where possible so that colonies could be checked for phase reversions, and contamination prior to sequencing however.

Consequently, for strain 0100993 non-reversible SCV's were sequenced, whereas for strain 37A, both SCV samples contained a mix of encapsulated and unencapsulated forms.

Samples were extracted using the phenol-chloroform method (see methods) due to the greater quality and quantity of DNA that could be extracted relative to automated methods available. Strains were then sequenced using a MiSeq platform, producing reads of approximately 150 bases.

6.2.6 *De-novo* sequence assembly

Control isolates were first assembled *de novo* using the tool VELVET. The contigs generated were ordered against closely related reference strains available from the NCBI using ABACAS. Annotations were subsequently transferred using the Rapid Annotation Transfer Tool (RATT). The resulting *de novo* reference assembly was then compared to the publicly available reference using ARTEMIS and ACT. This comparison was used to quality check the assembly, identifying whether gene order was conserved, and the degree to which the assembled reference differed from that of the publicly available reference.

Once a reference had been constructed, the sequence reads for all study isolates (i.e. including the reads for the *de novo* assembled reference) were remapped to this using SMALT and BWA (Figure 61). By comparing the mapping statistics and

BAM files produced by these read aligners it was possible to determine the quality of the reference assembly. In addition by re-mapping all 4 isolates to the control reference, it was also possible to determine whether inconsistences were present when comparing the re-mapped control to itself. Furthermore using BWA in addition to SMALT, it was possible to determine whether any differences arose due to the different alignment algorithms invoked in these programs, and this allowed indels to also be called by LASER. As indels are also called by the SMALT pipeline (Harris et al., 2013a), this allowed further verification of indels identified using both methods.

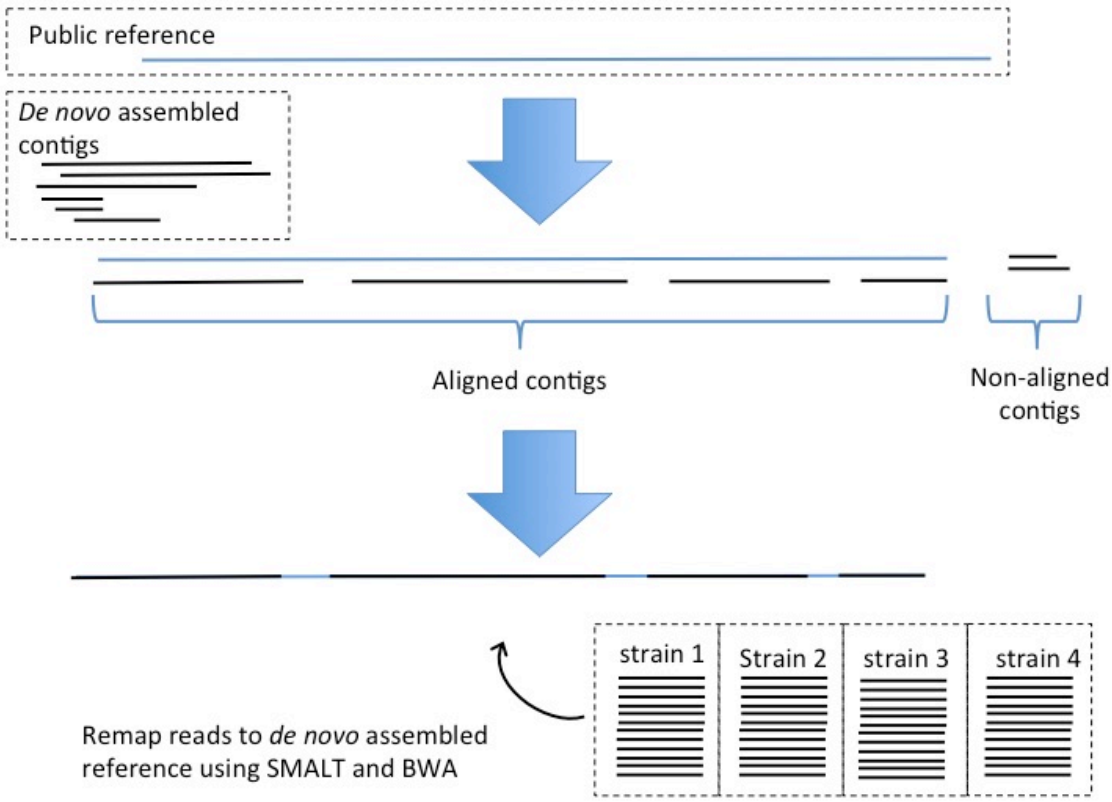


Figure 61: Following *de novo* assembly of control isolates, these contigs were aligned to a close reference sequence using ABACAS. The unaligned short reads for each isolate were then aligned to this reference, and mapping scores checked. Changing the public reference during contig ordering was found to affect the subsequent mapping quality. As such, the assembly with the highest mapping score was chosen for use in the subsequent analyses.

6.2.7 Assembly Quality

Using the above method differences in the quality of the mapping produced were identified, according to the public reference being used during contig ordering. Therefore the SMALT mapping was repeated for all closely related references

available from the EBI. The overall percentage of mapped reads across all of the sequenced isolates was then used in order to determine which public reference resulted in the best mapping.

6.2.8 010993

Using the percentage of reads mapped as a measure of assembly quality all *de novo* assembled references performed similarly well, with the exception of the reference created when using Spn034156 (Table 32). The assembly generated when using SPN994038 was chosen as this offered a fractionally higher overall percentage of reads mapped.

Reference used in contig ordering	Serotype	Control 1 (% mapped)	Control 2 (% mapped)
SPN034156	3	94.53697808	51.60064449
SPN034183	3	99.07489759	99.13506683
SPN994038	3	99.55134463	99.61434252
SPN994039	3	99.5464366	99.60861623
OXC141	3	99.05636214	99.12987495

Table 32: The overall percentage of reads mapped under SMALT according to the de novo assembled reference used for 0100993 isolates. The public reference used for contig ordering, and construction of the de novo reference used in the SMALT mapping is indicated.

The variation identified is summarised in Table 33. Interestingly a single base indel was identified upstream of the *spxB* gene within three of the four sequenced isolates. This region was identified as a possible promoter for the *spxB* gene using PEPPER (pepper.molgenrug.nl). This change involved the duplication of a guanine in mutant strains. The SNP was present in the two of the isolates generated, suggesting an ancestry between the isolates and the second control strain, rather than two independent occurrences of this SNP

Synonymous/Non-synonymous	Gene affected	Isolates	Type	Bases affected
Unknown	*spxB* promoter	Control 2, SCV1, SCV2	Duplication (G)	1
Frameshift	*argR*	SCV1, SCV2	Duplication (A)	1
Frameshift	SPN994038_09830	SCV1, SCV2	Deletion (T)	1
Inactivation (confirmed *in vitro*)	*cap3A* (ugd)	SCV1, SCV2	SNP	1

Table 33: Genetic variation identified in sequenced 0109933 isolates. The first column indicates the effect of the mutation- frameshifts are likely to inactivate the gene, although this was not confirmed *in vitro*. The second column indicates the gene containing the mutation. The third column indicates which isolates carried the mutation. In addition, the type of polymorphism, and number of bases affected is also indicated.

Both SCV isolates carried a single base duplication (adenine) in an *argR* gene, identified as part of the Arginine repressor family. The duplication present in both SCV's resulted in a frameshift, and indicated the possible inactivation of this gene. The arginine repressor appears to be responsible for the repression of the prokaryotic arginine dihydrolase pathway. This pathway is composed of three enzymes, which functions during anaerobic respiration to generate ATP through the break down of L-arginine, producing carbon dioxide and ammonia as additional by-products. Consequently such modification might be expected to effect anaerobic growth of these isolates.

A single base deletion (thymine) was found to occur in both SCV's. The affected gene, which again appeared to be inactivated by this frameshift encoded a putative DNA-binding protein and was part of the Thioesterase superfamily. The gene possessed a DNA binding domain as well as DRTGG, and CBS domains- these latter two domains frequently co-occurring. Whilst the function of the DRTGG domain is currently unknown, the CBS domain is thought to be involved in increasing the sensitivity of proteins to adenosyl carrying ligands (ww.ebi.ac.uk). Consequently, inactivation of this gene could affect the function of another protein, or proteins within the genome, with unknown consequences.

A non-synonymous, guanine to adenine SNP (cysteine to tyrosine) mutation was identified within the cap3A gene of both SCV isolates, with no other variation identified within the cap locus (Figure 62).

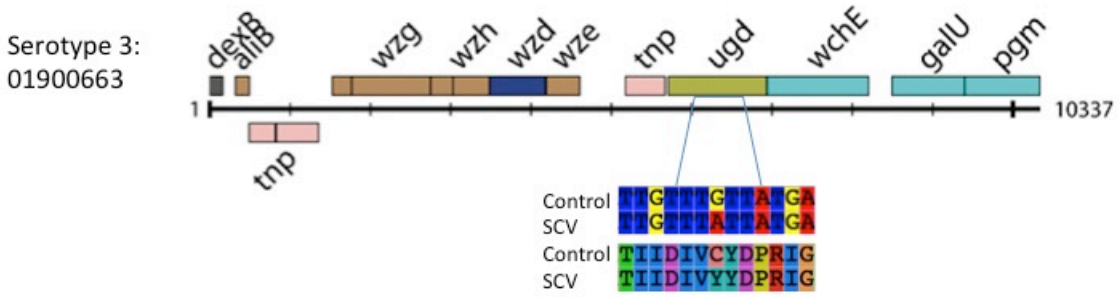


Figure 62: Serotype 3 capsule region. The nucleotide changes present between control and SCV isolates are indicated, relative to their position in the locus. The amino acid translation for the effected region is also shown.

To confirm the ability of this mutation to inactivate the *cap3A* gene the wild-type section of the *cap3A* gene was amplified from the control strains. In brief, the PCR cycles were reduced to 20 (from 35) and a high fidelity Taq polymerase used to ensure mutations were not introduced into the sequence. The PCR conditions were adapted from Waite and colleagues (2001). SCV's were then transformed by growing to an OD_{600} of 0.3, corresponding to the beginning of log growth. 1ml of culture was removed and 2µl of CSP peptide in addition to calcium chloride added, in order to promote competence. For isolates that regained capsule expression, the *cap3A* gene was amplified, purified through ethanol precipitation and capillary sequenced (Applied Biosystems 3730xl platform). The reads were visualised using gap4 to confirm that the SNP had been removed, and to check that the *cap3A* gene matched the control sequence (see methods for full details).

Using this approach it appeared that the substitution of a cysteine to a tyrosine at the amino acid position 239 was sufficient to inactivate this gene or render it unfit for purpose.

No other genetic variation was identified within these isolates.

6.2.9 37A

As mentioned previously, the 37A SCV's reverted at a high frequency and it was not possible to isolate non-reversible SCV's. Consequently a mix of SCV and reverted colony types were sequenced for each SCV isolate. *De novo* assemblies ran on each control indicated that control 1 assembled far more poorly, with 57 contigs generated compared to 14 in control 2. This appeared to be reflected in the mapping quality, with fewer reads mapping to the first control relative to the second control in most cases (Table 34). The reference aligned to SPN994038 was chosen as this gave the best percentage of reads mapping (Table 34).

Reference used in contig ordering	Serotype	control 1 (% mapped)	control 2 (% mapped)
SPN034156	3	58.73421856	54.1980321
SPN034183	3	87.01074845	94.55501942
SPN994038	3	87.46465632	94.56186423
SPN994039	3	87.42575917	94.55795129
OXC141	3	87.01505756	94.53937047

Table 34: The overall percentage of reads mapped under SMALT according to the *de novo* assembled reference used for 37A isolates. The public reference used for contig ordering, and construction of the *de novo* reference used in the SMALT mapping is indicated.

Some inconsistencies were identified between the different assembly methods in this analysis. The *de novo* assembly for control 1 contained a 2bp duplication and 45bp deletion when compared to all SMALT and BWA mappings. This implies variation can occur between *de novo* versus the mapping of reads to a predefined reference (SMALT and BWA). Interestingly, both of these mapping inconsistencies involved duplicated DNA, which appears to be an important problem for read depth based sequencing methods.

Furthermore, this 45bp deletion represented the removal of a duplication present in the *boxB* element. BOX elements are often highly conserved DNA sequences, occurring in three subunits (*boxA*, *boxB*, and *boxC*). *BoxB* duplications up to 4 repeats long were identified by Martin and colleagues (1992), and it is not unusual to identify multiple copies of this gene. Box regions are thought to be

involved in gene regulation, and have been found to be associated with competence and virulence encoding genes (Martin et al., 1992). In this case the BOX repeat region was flanked by the upstream occurrence of an *ilvD* gene, a gene thought to encode a DNA-binding protein. A second conserved hypothetical gene was located immediately downstream of the box region, with no information available as to its function.

A third duplication was identified within three of the isolates, but absent from control 1. This site was intergenic, and was not found to occur in the promoter region of a gene. However, when viewing the aligned reads over this site in control 1, a mix of reads was present, with 56% supporting the inclusion of this repeat (Figure 63).

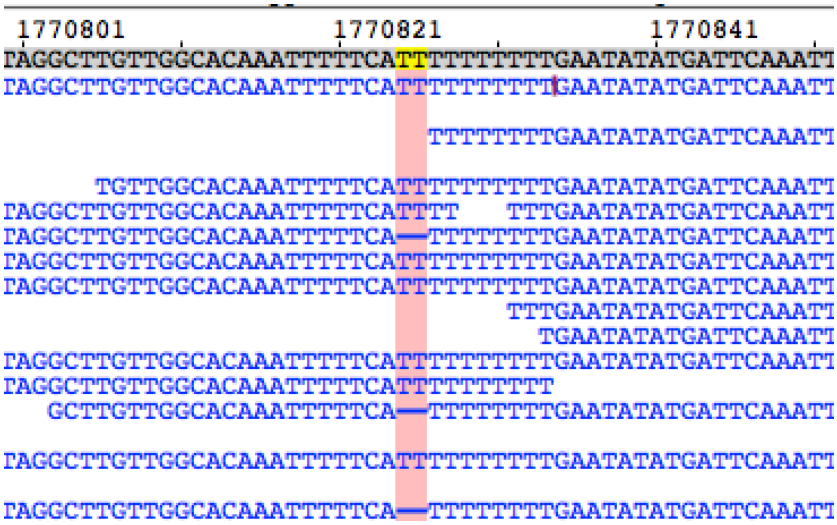


Figure 63: The BAM alignment, viewed in Artemis, indicated variation in the short reads mapped at this location. A subset of reads were found to consistently place a deletion at this location. However, mapping errors are often associated with repeat regions, and consequently in this case the variation most likely occurs due to sequencing errors. However, such variants were common, and it was not possible to determine to what extent such variation potentially occurs *in vitro*.

Interestingly a large number of these reads span the duplicated region, increasing support for this variant. This observation could indicate mixture within the sequenced population at this site, rather than a mapping issue. Multiple regions similar to this were similarly found, with a mixture of reads supporting the inclusion an indel, but were insufficiently supported to be recorded by LASER or PINDEL. No such tools however are currently available in order to identify and quantify the occurrences of these structural variants.

Consequently, this specific example is discussed in order illustrate a general finding, similarly identified among the other sequenced strains.

A final deletion was identified in SCV 1, which was found to be well supported in both SMALT and BWA mappings (Figure 64). This deletion measured 108bps long, and occurred within a gene labelled *pstI*, identified as encoding a phosphoenolpyruvate-protein phosphotransferase. This enzyme functions to catalyse the phosphorylsis of sugar substrates during their translocation across the cell membrane (www.ebi.ac.uk). Interestingly, although 108 of the 1736bp were removed during this deletion, the protein sequence remained in frame, suggesting the protein may remain functional despite this reasonably large deletion.

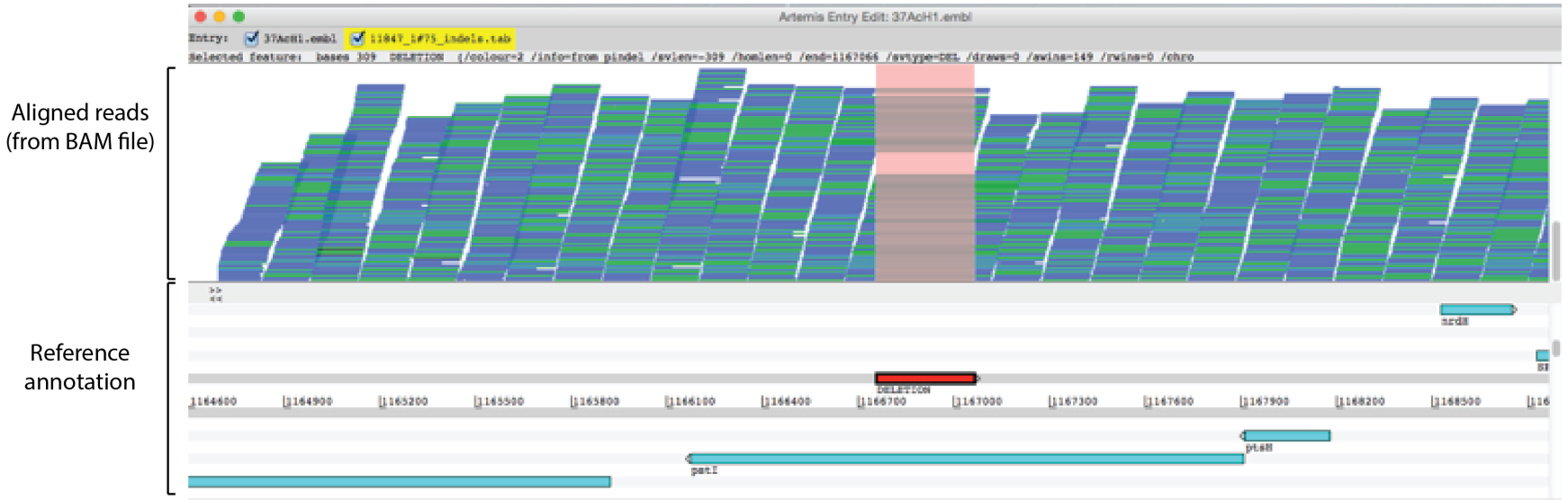


Figure 64: Location of the 109bp deletion identified within a PEP-phosphotransferase type protein. The BAM file is shows the alignment of mapped reads. Reference annotation indicates the location within the genome that is being viewed. Blue reads in the BAM alignment indicate reads present uniquely, whereas green reads indicate identical reads which have been concatenated to save space in the visualisation. The highlighted red section in the BAM filed indicates a large deletion, supported by all reads aligned at this location.

A SNP was identified in SCV 1, which resulted in the insertion of a premature stop codon within an ancestrally inactivated putative hyaluronate lyase, and as such would not be expected to have any effect on the gene's activity.

A second SNP was identified within SCV 2, a thymine to cysteine transition, resulting in a serine to proline replacement. This gene functions in the branched chain amino acid transport system, as part of the permease component (www.ebi.ac.uk), although it is unclear whether this SNP would have conferred any change in gene activity.

A third mutation was present in both SCV's, present within a gene identified as *trmE*, a tRNA modification GTPase. This gene functions to modify tRNA, introducing a 5-methylaminomethyl-2-thiuridine in the third position of some tRNAs (Prado et al., 2013). Once again the effect of this SNP on gene function was not investigated

A fourth SNP was introduced into the catabolite control protein A, *ccpA* gene of SCV 1. This adenine to thymine transversion resulted in an aspartic acid to valine amino acid substitution. The CcpA protein belongs to the LacI family of global transcriptional regulators, that are present in gram-positive bacteria. The gene functions to repress genes that utilise carbohydrate. Interestingly, growth rates in several bacteria have been found to be reduced following disruption of this gene (Luesink et al., 1998).

Two final mutations were present at adjacent intergenic locations, both flanked by the *mutS* gene upstream, and an ATP-transporter ATP-binding protein downstream. Both SCVs carried a cysteine to guanine transversion, whilst SCV 2 carried an additional adenine to thymine mutation adjacently. This region was not identified as a promoter for either gene, and so these SNPs appeared to be synonymous.

No disruption was identified within the capsule locus of 37A isolates, with the reads mapping consistently across this region for all isolates. This suggests that

although phenotypically variation would be expected to be present at this locus, the reads supporting this rearrangement were either discarded during the assembly process, or occurred at such a low frequency that they were not observable.

Synonymous/Non-synonymous	Gene affected	Isolates	Type	Bases affected
Intergenic (boxB)	n/a	control 1, control 2, SCV1, SCV2	Duplication	45
Intergenic	n/a	control 1, control 2, SCV1, SCV2	Deletion (repetitive)	2
Intergenic	n/a	control 2, SCV1, SCV2	Deletion (repetitive)	1
Unknown	*pstI*	SCV1	Deletion	108
Non-synonymous	CCP30406.1	SCV2	SNP	1
Synonynmous (in pseudogene)	pseudogene	SCV1	SNP	1
Non-synonymous	*trmE*	SCV1, SCV2	SNP	1
Non-synonymous	*ccpA*	SCV1	SNP	1
Synonymous	n/a	SCV1, SCV2	SNP	1
Synonymous	n/a	SCV1, SCV2	SNP	1

Table 35: Genetic variation identified in sequenced 37A strains. Duplication of the *boxB* gene, and intergenic 2bp deletion appeared to represent mapping and assembly issues as they were identified across all isolates when remapped to the *de novo* reference.

6.2.10 A42174

The assembly statistics for strain A42174 are presented in Table 36. *De novo* assembly resulted in both controls being assembled into 25 contigs, indicating a high level of overall read coverage. The reference aligned to SPNP1031 was chosen for the mapping analysis.

Reference in ABACAs	Serotype	control 1 (%mapped)	control 2 (% mapped)
P1031	1	91.27929084	94.3237561
INV104	1	58.81299198	44.83746205
SPN032672	1	66.8277789	52.70406969
SPN033038	1	66.77982063	52.63994047
PCS8235	1	49.30670806	23.86122655
SPN1041	1	91.21656414	94.34782696

Table 36: The overall percentage of reads mapped under SMALT according to the *de novo* assembled reference used for A42174 isolates. The public reference used for contig ordering, and construction of the *de novo* reference used in the SMALT mapping is indicated.

Genome-wide indel and tandem duplication variation was not detected in the A42174 study. This suggests that either these isolates are able to repair genomic damage at a higher rate than could be detected, or that they have a higher tolerance to hydrogen peroxide stress than isolates 0100993 and 37A. There was some evidence for the occurrence of indels when viewing the BAM files in ARTEMIS, but lacking sufficient read support these were excluded.

SNP identification indicated a non-synonymous cytosine to thymine transition was present in the *wchB* gene of both SCV isolates (Table 37). This resulted in the conversion of glutamine to a nonsense (or STOP) codon, likely inactivating the *wchB* gene. The product of the *wchB* gene is thought to work in conjunction with WchD, to assemble the repeated trisaccharide unit that composes the capsule (Aanensen et al., 2007). Inactivation of this gene is therefore likely to prevent the assembly of this structure, and lead to unencapsulation, as was expected given the colony morphology (Figure 65).

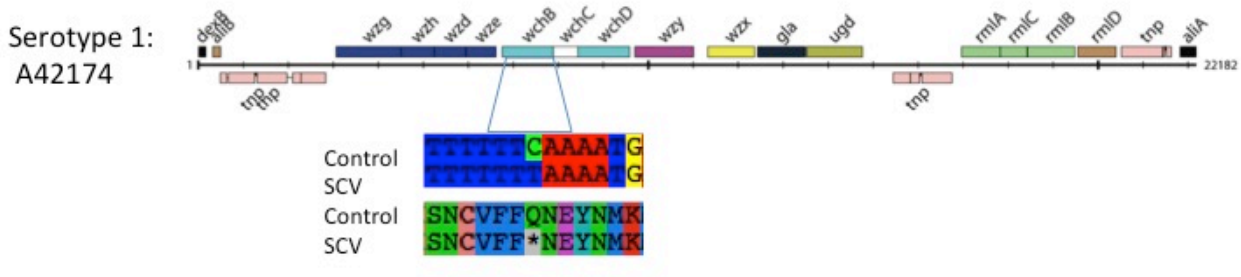


Figure 65: Serotype 1 capsule region. The nucleotide changes present between control and SCV isolates are indicated, relative to their position in the locus. The amino acid translation for the effected region is also shown.

A second non-synonymous mutation was identified in a *guaB* gene, encoding an inosine-5-monophophate dehydrogenase (Table 37). This enzyme is key to the regulation of guanine nucleotide biosynthesis, responsible for the rate-limiting step of this reaction (Collart and Huberman, 1988). Whilst this indicates a possible effect of the rate of this reaction, this was not investigated further.

Synonymous/Non-synonymous	Gene affected	Isolates	Type	Bases affected
Non-synonymous	*wchB*	SCV1, SCV2	STOP	1
Non-synonymous	*guaB*	SCV1	SNP (G->S)	1

Table 37: Genetic variation identified in sequenced A42174 strains.

6.3 Growth Curve Analysis

To identify the effect of hydrogen peroxide induced damage, control isolates and mutant SCVs were grown in batch culture to compare growth characteristics. Isolates were grown on plates, and prepared to an OD_{600} 0.03 through re-suspension of plated colonies in PBS. The bacterial suspension was immediately used to inoculate 198µl of BHI broth to give a final well volume of 200µl. Absorption readings were then recorded over a 16hr period (see methods). Repeats were made for each control and SCV, resulting in 8 technical repeats for each sequenced isolate.

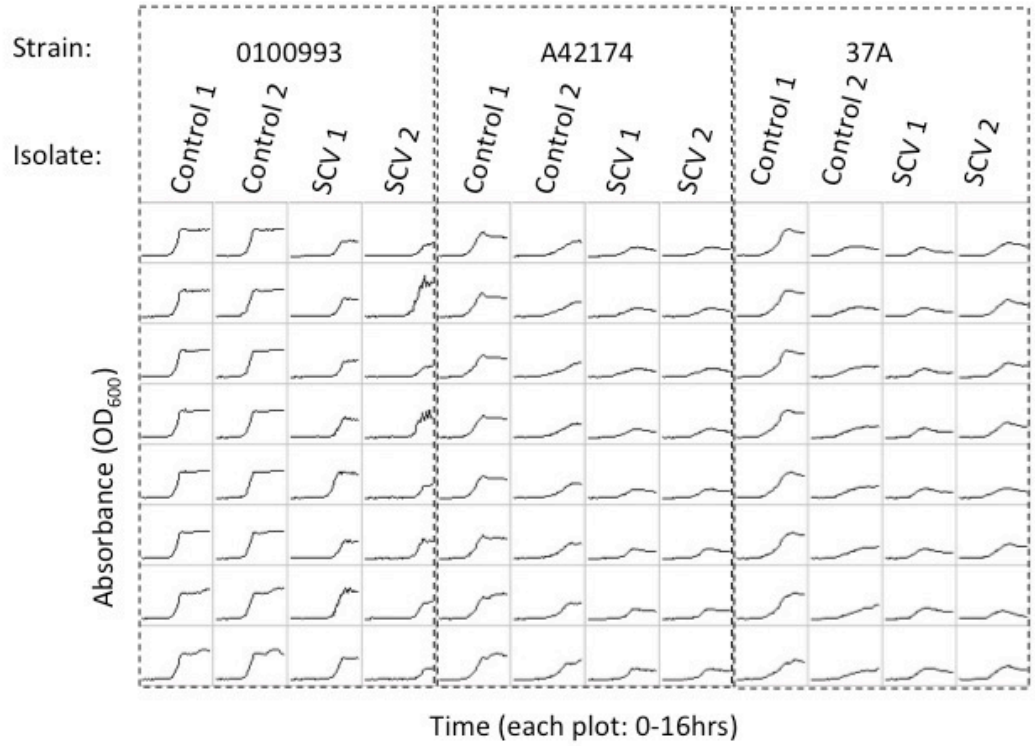


Figure 66: Growth curve analysis of control strains relative to the SCV isolates used in this study (measured at OD_{600}). Growth curves between control and SCV isolates of 0100993 showed consistent differences. However, the other isolates tested showed much greater variation. The lack of consistency suggests other factors than sample type may have been affecting growth.

Growth curves for strain 0100993 identified clear differences between the control isolates relative to the SCV isolates sequenced (Figure 66). Given the range of genetic polymorphisms identified, this suggests a severe growth impairment as a result of hydrogen peroxide induced damage. Interestingly doubling time, and the lag time showed the strongest difference between control and isolates. The final OD_{600} value achieved (stationary phase) however was much more variable. The observed trend could represent the random repair of stress-induced damage. This could result from the process of slipped-strand repair, which can lead to isolates repairing genomic damage at different frequencies (Waite et al., 2001).

In contrast there was far less consistency between the controls of strains A42174 and 37A. This suggests variation was present in the control population prior to our study. As such, it was not possible to identify whether additional growth impairment occurred in the SCV isolates relative to the control population. Given the observed growth curves for the SCV isolates did not fall

any lower than control 2 in both cases, this could indicate the maximum level of growth impairment that strains are able to tolerate before they are rendered unfit for survival.

The variation present in the control population suggests DNA damage is likely to have occurred following the routine aerobic growth of control populations prior to this study, restricting any further conclusion from being drawn from the study of isolates 37A and A42174.

As such the growth curves suggest genomic disruption may have been present, perhaps reflecting some of the genomic variation identified previously. However, given the inconsistencies between control isolates, there may have been other factors affecting the growth of isolates more reflective of environmental conditions than the underlying genetics. Consequently, the growth curve analysis was generally inconclusive, with only isolates of the type 0109933 growing consistently between controls and SCVs.

6.4 SpxB Variation Within Global Lineages

The SpxB gene is thought key to the production of hydrogen peroxide, therefore variation may effect the concentration of hydrogen peroxide produced, which could have implications for the ability of a pneumococcal strain to outcompete niche co-colonisers. Such an affect could be hypothesised to be evident in globally prevalent lineages, such as PMEN1 (section 1.3.8). To determine whether variation in the ability to produce hydrogen peroxide was evident within this lineage, a subgroup was identified that carried SNP polymorphism within the *spxB* gene and it's promoter (pepper.molgenrug.nl), as variation in the promoter region could effect *spxB* expression. The strain R6, an unencapsulated laboratory variant of D39, was used as a control strain. R6 lacks antibiotic resistance, and due to repeated laboratory culture, has undergone a degree of genomic degradation, as it has adapted to these conditions (Hoskins et al., 2001).

The *spxB* locus was highly conserved across the PMEN1 lineage, with only two strains out of 241 carrying mutations that could affect SpxB function (Figure 67).

Strain 08-B-120 carried a single mutation within the putative promoter, whereas strain MA28 carried a single non-synonymous mutation within the *spxB* CDS. Additional synonymous mutations were identified among 4 other strains.

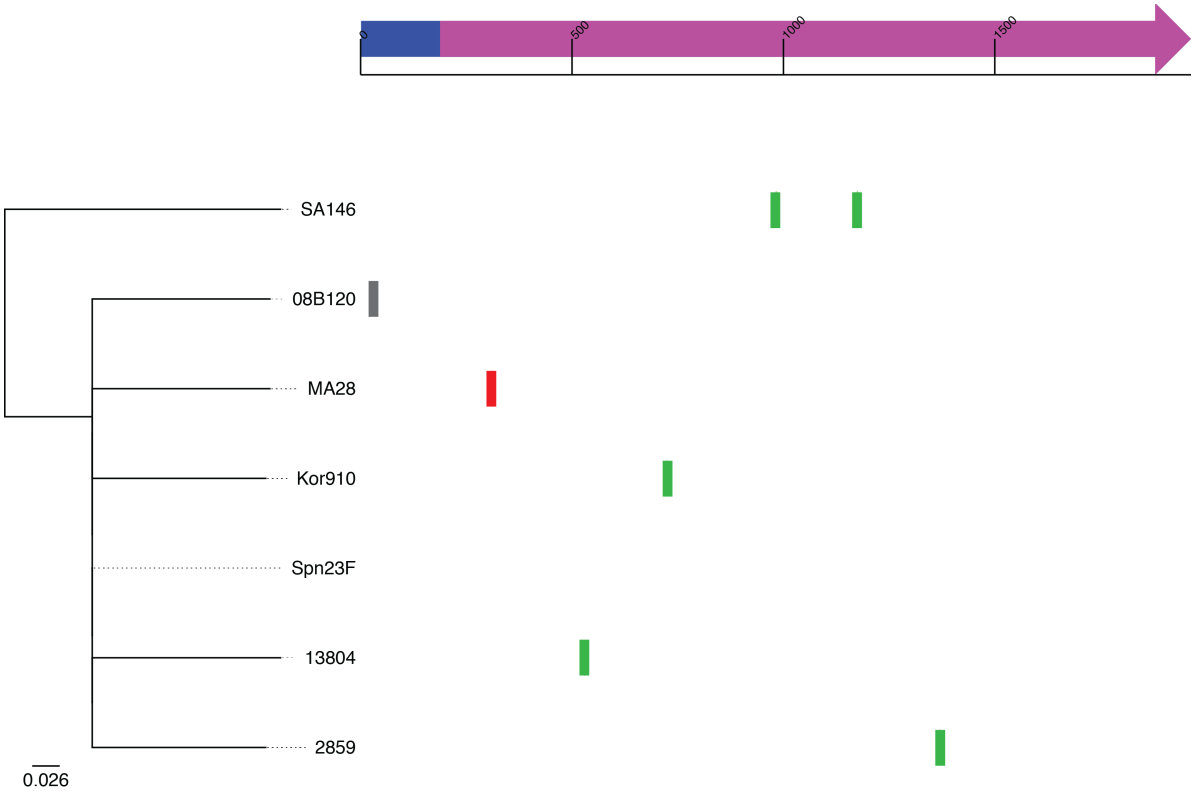


Figure 67: A phylogenetic tree constructed based on the nucleotide variation present in the spxB gene (pink) and its promoter sequence (blue). Non-synonymous SNPs within the CDS are coloured red, whereas synonymous changes are coloured blue. A single SNP (grey) was identified in the promoter region of strain 08B120.

Based on the nucleotide polymorphism identified in the phylogenetic analysis (Figure 67), absorbance was measured using the Amplex® Red assay to determine variation in the ability for these strains to produce hydrogen peroxide. In addition the R6 was used as a comparator (Table 38, Figure 68).

Strain	Origin	Serotype	Other names
R6	Acapsular D39 variant	2 (historically)	R6 variant
2859	Germany	23F	PMEN1
13804	Germany	23F	PMEN1
Kor910	Korea	23F	PMEN1
SA146	South Africa	23F	PMEN1
Ma28	Turkey	23F	PMEN1
08-B-120	Korea	23F	PMEN1

Table 38: Isolates used to determine variation in hydrogen peroxide produced after 8 hours of growth in broth cultures.

In terms of overall hydrogen peroxide production, all PMEN1 isolates appeared to produce similar concentrations. Despite the presence of SNPs within 08-B-120 and MA28, no clear correlation with hydrogen peroxide concentration was identified.

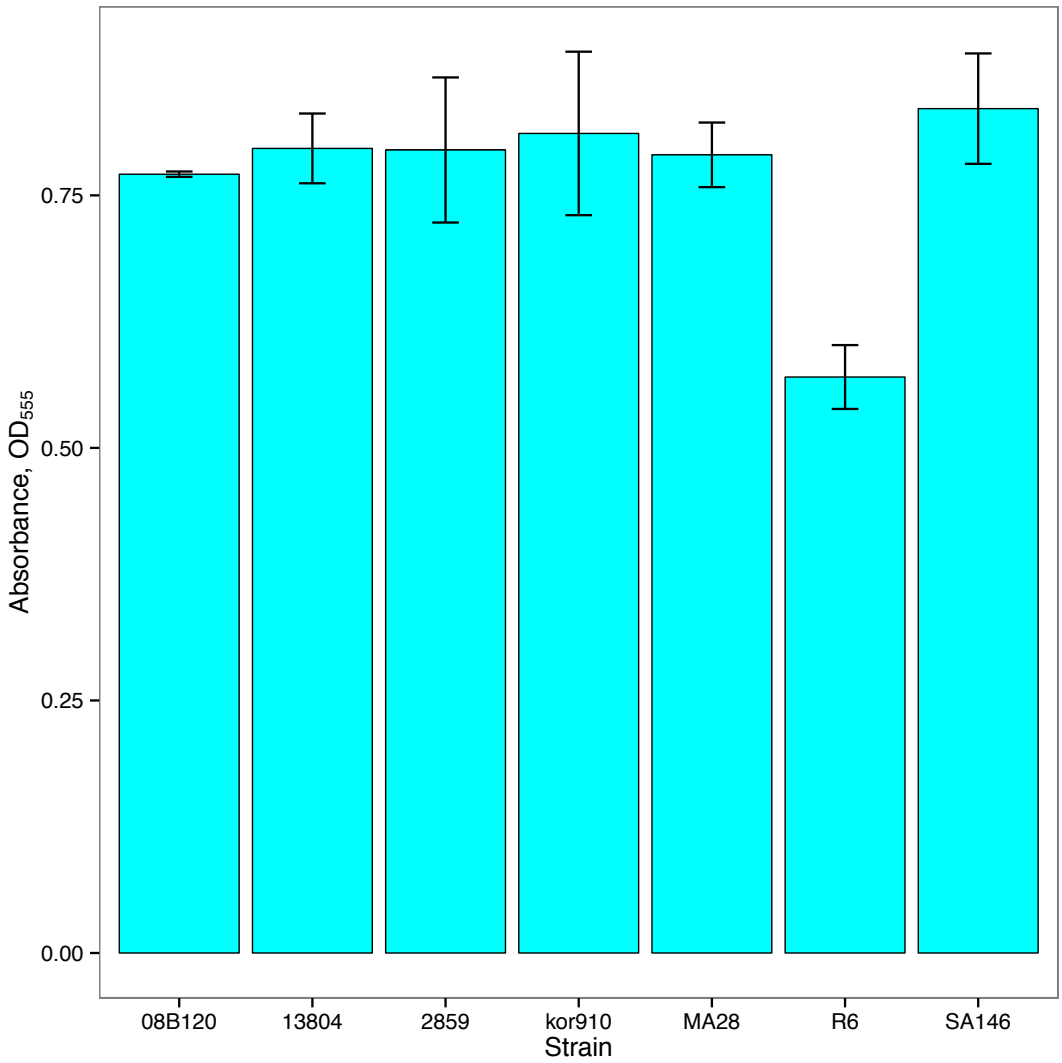


Figure 68: Raw absorbance values for PMEN1 isolates compared to the control R6 variant, indicating less hydrogen peroxide was produced by the R6 isolate compared to the PMEN1 isolates. Despite the genetic variation identified in the spxB gene and its promoter region, PMEN1 isolates appeared relatively consistent in their ability to produce hydrogen peroxide.

Overall PMEN1 isolates were found to produce substantially more hydrogen peroxide than R6 (0.80mM/mL compared to 0.57mM/mL). However, there was little indication that the variation identified in the *spxB* locus affected hydrogen peroxide production in these isolates. Instead, this locus was highly conserved across the 241 PMEN1 isolates tested. Consequently, it may be important to measure hydrogen peroxide production between lineages.

6.5 Discussion

This chapter describes the ability of exogenously administered hydrogen peroxide to induce genome-wide damage within different pneumococcal strains. In addition an attempt was made to correlate variation in the activity of the *spxB* gene with the differing ability of isolates to produce hydrogen peroxide within an important MDR pneumococcal clone.

6.5.1 Genetic Damage

Whilst the ability to cause genetic damage by hydrogen peroxide is well documented, this is the first study to avail of the power of high-throughput sequencing to attempt to identify the genome-wide effects of this process. Nasopharyngeal colonisation was modelled using the sorbarod growth method, and colonies were subjected to 2mM concentration of hydrogen peroxide, delivered in the media over a 48hr period. Using three clinically important strains, SNP, deletion, and tandem duplication differences were identified between controls and SCV isolates, and within control and SCV populations. The observed genome-wide disruption is consistent with hydrogen peroxide induced damage (Pericone et al., 2002). However, the observed damage was lower than would be expected, consistent with the pneumococcus being able to maintain genomic integrity despite exposure to high levels of oxidative stress.

For strains 0100993 and A42174 it was possible to identify non-reversible SCV, whereas this was not possible for strain 37A, with SCVs reverting at a high frequencies for this isolate. To test the ability of current sequencing platforms to process mixed reads, SCVs for the 37A stain were sequenced. Such populations contained a mix of encapsulated and unencapsulated colonies, and as such variation in the capsule synthesis genes would be expected. Despite this, no variation could be identified at these loci *in silico.* This indicates that some levels of sequence diversity cannot be detected using current technologies, which require large amounts of starting material, in addition to high read depth and support for the accurate detection of genomic variation. As such, the assemblies resulting from such technologies represent a consensus sequence, whereas

additional variation may still be present in the original sample. Indicated in this study, in some cases large numbers of reads were identified that provided consistent support for an indel. However, the current cut-off criteria, meant that such events were disregarded. Such occurrences could indicate mixture within the sequenced population, or conversely, may represent sequencing errors, such as due to dephasation (as discussed in the introduction). Further investigation into the true nature of such events may therefore be important for improving this technology, and the interpretation of read data in the future.

6.5.2 Comparison with the studies of Waite

Sorbarod static growth can support pneumococcal growth for at least 27 days, with the proportion of SCVs isolated from the biofilm increasing over time (Waite, 2001). Such long-term growth indicates the sorbarod method is a good model for the extended pneumococcal growth periods that occur during *in vivo* nasopharyngeal colonisation, which can occur for periods of over 6 months (Gray et al., 1980). The sorbarod growth model can be viewed as a selector of small colony variants, which increase in proportion over time under aerobic conditions (Waite, 2001). Capsule expression typically covers surface adhesions (such as PsaA, and CbpA), which are targeted by the hosts immune system. Consequently, capsule expression helps subjugate the hosts defences. During *in vivo* host colonisation, down-regulation of the capsule is important in order to allow the interaction between pneumococcal adhesins and epithelial surface receptors. Waite (Waite, 2001) has previously suggested that capsule loss in the sorbarod system may therefore allow these adhesins to bind to the cellulose sorabord fibres. However, capsule expression also represents an energetically demanding task. Consequently, loss of the capsule may favour more rapid colonisation of this niche, relative to competing encapsulated forms, which must direct energy and resources to apparently superfluous capsule biosynthesis. The energetic costs involved in capsule production may therefore offer a more plausible explanation as to the selective advantage of capsule loss during sorbarod growth. This hypothesis is also supported by the finding that unencapsulation often follows extended periods of lab culture in pneumococci

(Hoskins et al., 2001). As such, whilst capsule loss may be favoured both during nasopharyngeal colonisation and during growth on the sorbarod, the drivers of unencapsulation are likely quite different i.e. to expose surface adhesion molecules, versus reducing resource consumption).

During Waite's (Waite, 2001) original sorbarod studies anaerobic and aerobic growth was compared. Interestingly Waite suggests anaerobic growth strongly favoured unencapsulation. Here, under anaerobic growth 100% of the colonies sampled from the effluent were found to be unencapsulated- with both phase-variable and non-reversible SCV types identified after only 1.5 days of growth. In comparison, under aerobic conditions, even after 24 days of growth, encapsulated colonies were still identified in the supernatant samples. Unfortunately Waite did not repeat this observation across different strains, with only strain 0100993 directly compared.

This suggested that SCVs may exist normally in the population, but occur at such low frequencies that they are generally not observed. Interestingly, capsule loss can occur after extended periods of lab culture (Hoskins et al., 2001). This is likely because the capsule is no longer required under laboratory growth, however, capsule loss does not emerge rapidly. The reason unencapsulated forms do not emerge more rapidly during plate culture perhaps indicates that there is not a high selective pressure to reduce energy costs under such growth. In addition, researchers will deliberately pick typical, rather than atypical colonies for subsequent culture. In contrast, sorbarod growth appears to favour capsule loss. This effect is perhaps enhanced under anaerobic growth, which being less efficient than aerobic respiration, increases the selection for the more efficient unencapsulated forms.

6.5.3 Inter-Strain Variation in Colony Characteristics

Strain 0100993 produced the most consistent results, with a clear difference between SCVs and controls during spectrophotomer growth analysis, and the ready isolation of non-reversible SCVs. In comparison, strain 37A behaved

markedly different, despite the both isolates belonging to the same ST. SCVs isolated from the 37A population reverted at a high frequency, and it was not possible to isolate non-reversible forms. As such tandem duplication changes were the predominant driver of capsule loss in this group (Waite et al., 2001), and appeared to occur at a different frequency than for strain 0100993. This trend occurred despite repeated growth of the 37A strain in the sorbarod apparatus. As such there appeared to be an underlying genetic difference between these populations, leading to differences in the ability for unencapsulation to occur.

6.5.4 The Biological Implications of DNA Damage and Hydrogen Peroxide Production

Although isolates were exposed to high levels of exogenous hydrogen peroxide, very little genetic variation could be identified between control and SCV isolates. Spectrophotometer analysis suggested a consistent difference in the growth characteristics between controls and SCV isolates of 0100993 strains. Whilst this could have reflected the genetic variation identified between control and SCV populations, no consistent differences were identified when similar growth analyses were conducted for A42174 and 37A strains. As such, this analysis was inconclusive in terms of identifying whether growth impairment had resulted from previous exposure to hydrogen peroxide in the sorbarod system.

Furthermore, much of the damage identified *in silico* was related to tandem duplication changes, or SNPs, which could equally reflect replication errors occurring independent of exposure to exogenous hydrogen peroxide (Waite et al., 2001, Waite et al., 2003). The lack of wider damage directly linked to oxidative stress could be due to the high reactivity of the hydroxyl radical, though responsible for most genetic damage (Cooke et al., 2003). This molecule has a very short diffusion ability, perhaps due to its high reactivity. Consequently, such radicals must form close to genetic material in order to induce damage (Greinert et al., 2012, Redmond and Kochevar, 2006). Although hydrogen peroxide was present in the media, it would therefore have to enter

the pneumococcal cells, before breaking down to yield hydroxyl radicals, close to the DNA. Hydrogen peroxide is found to readily diffuse through cell membranes, however the pneumococcus is known to possess mechanisms to protect against such exogenously derived hydrogen peroxide (McAllister et al., 2004, Andisi et al., 2012). Therefore, whilst endogenous hydrogen peroxide clearly has the ability to inhibit pneumococcal growth (section 6.1), the external application of this within the media may not be an informative model for hydrogen peroxide induced stress occurring *in vivo.*

6.5.5 SpxB and Hydrogen Peroxide Production in PMEN1

Differences in hydrogen peroxide production between pneumococcal strains could lead to differences in *in vivo* colonisation proficiency. Higher concentrations of peroxide might be expected to aid the clearance of nasopharyngeal co-colonisers, aiding the establishment of the host strain (Regev-Yochay et al., 2007). Isolates from the highly successful PMEN1 lineage might therefore be expected to produce high concentrations of hydrogen peroxide, perhaps helping to explain the success of this clone. In order to investigate whether isolates within this lineage varied in their ability to produce hydrogen peroxide, genetic variation with in the *spxB* gene and its promoter region was assessed. This indicated that the PMEN1 isolates studied here possessed a high level of sequence conservation at this site, indicating that hydrogen peroxide production, resulting from the action of the *SpxB* gene (Carvalho et al., 2013), was likely consistent across the lineage.

Only two isolates were found to carry non-synonymous polymorphisms within this locus, out of a dataset of 241 PMEN1 isolates. However, there was no indication that these mutations lead to any difference in the ability of these two isolates to produce hydrogen peroxide. Interestingly, all PMEN1 isolates were found to produce more hydrogen peroxide than the D39 isolate tested. This suggests that an expansion of this analysis to include other pneumococcal lineages may be informative. For instance, this include a comparison between

PMEN, and non-PMEN lineages, in order to identify whether any consistent differences exists between these populations.

6.5.6 Sequencing Variation

This study has highlighted some of the difficulties current sequencing platforms have in identifying genomic variability. The reference sequence used during the ordering of *de novo* assembled contigs was found to affect the overall mapping quality. Here, variation was identified between the remapped reference, and the *de novo* assembled contigs. The remapped sequence appeared to offer the more reliable sequence in such cases. Such variation was also frequently associated with repeat regions, which therefore still appear to be problematic for assembly algorithms, despite the increases in read length that have occurred since this technology was first developed.

A number of cases where also identified where a subset of the mapped reads indicated a deletion was present. However, due to the mapping thresholds in place, such indels were excluded. Whilst such events could reflect mapping errors, in some cases such events were found to be consistently located within the middle a large proportion of the mapped reads present. If such events purely resulted from errors occurring during the mapping stage, it might be expected that these events would be found more randomly places within the mapped reads, and more often at the start of end position of these reads. A second possibility is that such events represent sequencing, rather than *in silico* assembly errors. Dephasing describes the tendency for reads to differ in length as sequencing progresses. This occurs when a polymerase fails to add a base on the read being sequenced, or incorporates too many bases in one go. As sequencing occurs in dense clusters, random polymerase errors such as these, cause reads to gradually fall out of sync as the sequencing process continues. Consequently, the sequenced dephased-read will be shorter or longer, leading to the false deletion or insertion of sequence. Such events are therefore often associated with repeat regions of sequence. Determining the true nature of such

polymorphisms may therefore be important for the better interpretation of sequence data.

6.6 Concluding Remarks

Oxidative stress is thought to result in widespread genomic damage, through deletion, tandem duplication and single base mutation in pneumococci (Pericone et al., 2002). Here an attempt was made to assess the ability of exogenously administered hydrogen peroxide to cause such damage, availing of modern high-throughput sequencing to identify such genetic polymorphisms. To better model the extended pneumococcal carriage, that occurs following colonization of the nasopharynx, study isolates were cultured using a sorbarod apparatus, as has been described previously (Waite, 2001). Despite culturing strains for 48 hours under continuous exposure to toxic concentrations of hydrogen peroxide, very little genetic damage was identified. Furthermore, much of this variation could equally have been attributable to endogenously produced hydrogen peroxide, or sequencing errors (Pericone et al., 2002). The pneumococcus therefore appears to be highly competent at subjugating the pressures of hydrogen peroxide when administered in this way.

Oxidative stress may therefore occur differently in the clinical environment. Such exogenous sources of oxidative stress include exposure to neutrophils, macrophages and additional lactic acid bacteria present in the nasopharynx (Andisi et al., 2012). Therefore, perhaps radical attack is better targeted at invading pneumococci in such situations. As such it may be more informative to study pneumococci subject to high levels of oxidative stress clinically (Carter et al., 2014), rather than modeling this *in vitro.*

7 Recombinational Repair in *Streptococcus pneumoniae*

7.1 Introduction

Sources of genetic damage are ubiquitous, yet maintaining a level of sequence synteny is essential for the long-term viability of all DNA based organisms. DNA damage can result from both internal functions, such as replication errors, to exposure to a wide variety of genotoxic stressors (Baharoglu and Mazel, 2014). Free radical attack is a particularly prevalent cause of genetic damage, which has recently been linked to a number of apparently unrelated causative agents, such as antibiotic therapy, and exposure to UV light (Kohanski et al., 2007). For the pneumococcus, incomplete pyruvate metabolism results in the endogenous production of hydrogen peroxide. Such molecules can act as a source of free radicals within the cell. Hydrogen peroxide can yield reactive oxygen singlets as well as the hydroxyl free radical. Whilst neither hydrogen peroxide nor oxygen singlets are thought to directly react with DNA, the hydroxyl free radical is particularly genotoxic (Figure 69)(Halliwell and Aruoma, 1991, Kawaguchi et al., 1996, Dizdaroglu et al., 2002). Perhaps due to its high reactivity the hydroxyl free radical has a very limited diffusibility, and consequently, must form close to genetic material in order to cause damage (Greinert et al., 2012, Redmond and Kochevar, 2006). In addition, the pneumococcus is exposed to a number of exogenous sources of free radicals during colonisation of the nasopharynx (Andisi et al., 2012). Whilst the pneumococcus has a number of mechanisms to protect against externally (Andisi et al., 2012) and internally derived hydrogen peroxide (Pericone et al., 2003), it is still susceptible to oxidative stress (Pericone et al., 2002). Once inside the cell hydrogen peroxide can react with free iron forming hydroxyl free radicals via the Fenton reaction. Hydroxyl radicals can in turn react with nearby deoxyribose bases causing a wide variety of DNA lesions. This can include disruption of methylation patterns affecting protein expression (Franco et al., 2008). In other cases, extraction of hydrogen atoms from nearby deoxyribose sugars leads to the formation of reactive sugar radicals, which subsequently react leading to the release of purine and pyrimidine bases. The resultant single strand and double strand breaks (SSB and DSB respectively)

can be lethal to the cell (Halliwell and Aruoma, 1991, Dizdaroglu et al., 2002, Kohanski et al., 2007, Dharmadhikari et al., 2014, Greinert et al., 2012).

It is essential that organisms protect and prevent against such damage. Consequently a number of highly conserved and integrated systems are invoked in order to detect, repair, and minimise the opportunity for damage to occur. Such systems include both error-prone, and error free repair mechanisms such as base excision, mismatch repair, and nucleotide excision (Sancar and Reardon, 2004, Friedberg, 2006, Iyer et al., 2006). Many of these mechanisms are invoked irrespective of whether the damage results from oxidative stress, or due to replication errors (Baharoglu and Mazel, 2014). First described by Radman (1975) in *E. coli*, the SOS response has become a model for DNA repair among bacteria. In its classical form, the system is activated when presynaptic RecBCD or RecFOR complexes detect single-stranded DNA (ssDNA). Following its recruitment, RecA binds to the ssDNA in the form of a nucleofilament, catalysing the autoproteolytic release of the LexA repressor from promoter regions of the SOS regulon, resulting in SOS induction (Baharoglu and Mazel, 2014).

In *E. coli,* the SOS response is associated with the expression of over 1000 genes, although only a fraction of these appear to be under direct LexA repression (Fernandez De Henestrosa et al., 2000). Of particular importance is the induction of competence, and the activation of error prone polymerases which help to repair genetic damage directly. Error prone polymerases function to repair damaged DNA that the normal proofreading proficient DNA polymerase is unable to act on. However, these polymerases lack a proofreading mechanism, and are poorly processive (Erill et al., 2007). Consequently, whilst these polymerases help repair DNA, they lead to the introduction of mutations in doing so.

SSBs that fail to be repaired by these mechanisms are converted to DSB, which may be repaired by homologous recombination. RecN appears to play a role in the initial recognition and recruitment of recombination proteins to the damaged DNA (Alonso et al., 2013). End processing then occurs, during which aberrant

bases are removed by resection, leading to the formation of 3'-tailed duplex DNA. In *E. coli* this process appears to be moderated by a conserved octanucleotide sequence, termed "Chi" sites (Kulkarni and Julin, 2004). Here, recruitment of RecBCD to the broken DNA leads to concurrent exonuclease-helicase activities until the Chi site is reached, whereupon the RecBCD exonuclease activity on the 3' strand is suppressed leading to the formation of the ssDNA tail (Quiberoni et al., 2001, Dillingham and Kowalczykowski, 2008). SsbA binding then protects the 3' ssDNA tails on either side of the breakpoint, whilst tethering of DSB ends by RecN aids localisation, or "DSB coordination" and the formation of a repair centre (Alonso et al., 2013). Loading of RecA onto one of the ssDNA ends at the repair centre, promotes strand invasion by intact homologous genetic material, resulting in a displacement loop (D-loop) recombination intermediate. DNA synthesis then completes the process, being primed off the invading strand (Alonso et al., 2013).

In cases where homologous recombination is impaired, or template DNA is not available for repair, non-homologous end-joining (NHEJ) occurs, ligating the two sheared DNA ends together, however this results in the loss of the genetic material located between the break points (Ayora et al., 2011). This ligation process has been described in several species, mediated by Ku and ligase proteins (Bowater and Doherty, 2006, de Vega, 2013).

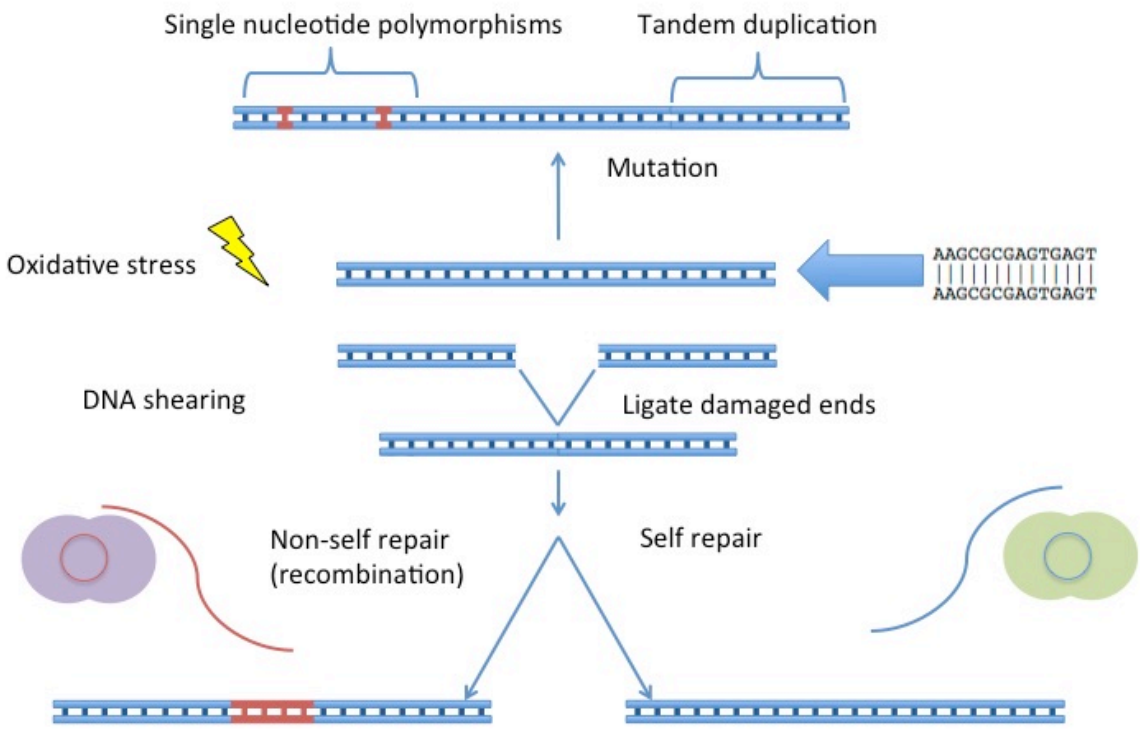


Figure 69: Diagram showing the types of genetic damage that are hypothesised to occur as a result of exposure to the hydroxyl radical. Where double strand breaks occur, recombination is required in order to repair such damage. Two mechanisms of repair were hypothesised to occur: either the isolates is transformed with material from a divergent population (non-self), or it recombines with its siblings, with no nucleotide variation being introduced. Recombination has previously only been detected in cases where divergent clusters of SNPs can be identified (Croucher et al., 2015).

Variations in the SOS response and repair mechanisms have been identified widely among prokaryotes (Erill et al., 2007). The potential for genetic disruption from the SOS response means that this mechanism must be closely regulated during occasions when it is not needed (Baharoglu and Metzel 2014). As such the SOS response is often tailored to the specific niche requirements of the host. Of the 33 genes under LexA repression in *Bacillus subtilis*, a gram-positive bacterium, homologues of only 8 are present in *E. coli* (Au et al., 2005). Similarly the system seems to have been reduced to just 15 genes in *Pseudomonas aeruginosa* (Cirz et al., 2007) compared to ~40 genes under direct LexA repression in *E. coli* (Fernandez De Henestrosa et al., 2000). In even more extreme cases, despite the apparent importance of the SOS system, it is seemingly absent.

The pneumococcus although possessing a RecA homolog, lacks LexA, and the error prone polymerases that form a fundamental part of the SOS response. As such, it does not appear to possess a functional SOS system. Despite this, the pneumococcus occupies an aerated niche, is exposed to immune attack, antibiotic therapies, and produces millimolar concentrations of hydrogen peroxide as a result of its metabolism (Pericone et al., 2000). It furthermore lacks catalase, OxyR and PerR regulons, and other components commonly used to protect against oxidative stress (see section 6.1)(Gennaris and Collet, 2013). Whilst the pneumococcus does possess other protective mechanisms, endogenous hydrogen peroxide production is found to promote genetic instability in the pneumococcus (Pericone et al., 2002, Prudhomme et al., 2006). Lacking the error prone polymerases of *E. coli,* it has been suggested that competence could play an important role in the repair of such damage, which is largely expected to results from the pneumococci's own metabolism (Auzat et al., 1999, Pericone et al., 2002).

However, a reliance on recombination in order to repair such damage would be expected to drive rapid diversification across the species, yet, the pneumococcus possesses a remarkably high level of sequence synteny within its housekeeping genes, at 1.1% diversity (Hanage et al., 2006). The MLST diversity for its coinhabitors and closest relatives are estimated to be considerable higher: *S. pseudopneumoniae*, 3.0%, *S. mitis*, 5.1%, and *S. oralis*, 6.2% (Hanage et al., 2006). Recent estimates of the likelihood of a SNP being introduced by recombination rather than mutation are also considerable lower at 7.2 (Croucher et al., 2011) compared to the previously estimated ~66 (Feil et al., 2000). There is therefore a paradox in terms of the internal and exogenous drivers for sequence diversity, relative to the high levels of sequence synteny observed within this species.

7.1.1 Study Aims

The aim of this study was to develop a method for the reliable detection of deletion and insertions, which could be applied to whole genome sequences.

Using this approach, the frequency and mechanisms of deletion repair were then investigated in an attempt to determine the potential role recombination has in the repair of genetic damage.

7.2 Method Development

The availability of whole genome sequences has increased exponentially since the arrival of NGS (Read and Massey, 2014), allowing datasets of 1000s of bacterial genomes to be analysed in parallel (Chewapreecha et al., 2014a). Such large-scale analyses have been made possible through reference-based assembly. A high fidelity reference is typically constructed using the Roche/454 platform, which typically produces read lengths between 200 and 400bp (Metzker, 2010). *De novo* sequence assembly can then be used to construct a draft genome, composed of a series of contigs. PCR primers targeted at the gaps between contigs, and subsequent sequencing of these products can be used to construct a high quality reference (Croucher et al., 2009). Faster and cheaper sequencing technologies are often then employed to expedite the sequencing process, generating gigabytes of data, but with reduced read lengths. By aligning these shorter reads, typically between 30-100 bases in length, to a pre-defined reference sequence, potentially many thousands of genomes can be assembled relatively quickly and cheaply.

The computational task of identifying all of the possible locations that a single read can map to in a genome of potentially $3.3x10^9$bp is highly resource demanding. To accelerate this process, read assembly programs are designed to first reduce the search space, before carrying out the more computationally demanding task of aligning each read to each possible location, and scoring each of these alignments. In order to assign a score to an alignment the assembly program must take into consideration the number of polymorphisms between the aligned reads, the accuracy in the base calls given by the sequencer, and the possibility of gaps. By assigning a cost to each of these potential variants, the "best match" for the position of each read compared to the reference genome is determined. The problem of identifying gaps or insertions of sequence is

particularly complex, as read assemblers must align sequence reads individually. Consequently, assembly programs are often limited in their ability to detect longer indels (>50bp) (Marschall et al., 2013).

Analyses to date have predominantly focussed on the more readily detectable SNP variation, which has formed the basis for most genetic studies to date- such as phylogenetic analysis, recombination detection, and GWAS studies. This is partly due to the difficulties in accurately detecting indels, and a consequent focus of comparison of the consensus sequence between isolates. However, recent advances in sequencing technology and analysis software have meant that although read length remains a barrier to accurate indel calling, a number of reliable indel detection tools are now available (Marschall et al., 2013).

Several approaches have been developed for identifying indels, which typically rely on the input reads comprising paired-end rather than single end sequences (section 1.7.3). *De novo* assemblers, such as Velvet (Zerbino and Birney, 2008) allow the detection of indels is discussed above. However, the issues associated read alignment in these programs has lead to the further development of re-aligners, such as the Genome Analysis Toolkit (GATK) (McKenna et al., 2010, DePristo et al., 2011). Such programs identify indels once the initial mapping has taken place. In addition, programs that identify differences in distances between two mapped paired end reads (e.g. BreakDancer, (Chen et al., 2009), and CLEVER (Marschall et al., 2012)), or assess differences in read coverage throughout the genome (Medvedev et al., 2009, Abyzov et al., 2011, Marschall et al., 2013) have also been developed.

Split-read alignment offers another approach to indel detetion. Split read aligners are focussed on detecting indel breakpoints that occur within the length of a single read. In cases where only the beginning or end part of a read maps to the reference sequence, this may indicate the presence of an indel. Pindel (Ye et al., 2009), and LASER (Marschall, 2013) are examples of such split read aligners. In cases where the start and end of a read can map by splitting the read, this suggests a deletion has occurred. Whereas in cases where both ends of a read

map adjacently when an internal portion of the read is excised, this suggests a sequence insertion has occurred. However, one limitation of this approach is that as a split read aligner infers the position of an indel based on a single read, when an insertion size exceeds the read length, it will not be detected, as only one end will map (Marschall et al., 2013). In contrast, for deletions, as the read should contain information about the start and end position of this event, this upper limit on detection is theoretically removed.

The method used to identify indels in this study was based on the assembly method developed at the Sanger Institute, and has been described previously (Harris et al., 2013a). By default this method uses SMALT to align reads in conjunction with the split-read aligner Pindel (Ye et al., 2009) for indel detection. Using paired-end reads, Pindel detects indels by first identifying reads in which only one of the read pairs has mapped. Additionally, this read must map uniquely in the genome, with no mismatch bases, and the unmapped read must not be mapped in the genome under the default alignment score threshold (in this case a Phred score of 30). Pindel uses this mapped read to determine the anchor point and direction of the unmapped read. The unmapped read is then compared to the reference using this information, allowing the search space for identifying possible alignments to be reduced. The search space for deletions is user defined, whereas the search space for insertions is the length of the unmapped read minus 1 (i.e. the minimum size of an insertion being 1bp). In cases of a deletion the read will be split into two, whereas insertions require the unmapped read to be split into three sections.

Pindel uses the aligned read of the pair to determine the direction of the unmapped read, and in doing so, the 3' and 5' ends of this read. Using a pattern growth approach, Pindel then attempts to identify minimum and maximum unique substrings that occur between the unmapped read and the genome sequence, for example:
The pattern "ATGCA" represents the unmapped read, and the sequence "ATCAAGTATGCTTAGC" represents the genome. The alignment algorithm begins searching for a match from the left-hand side of the pattern, i.e. the base "A". This

occurs five times in the sequence. Continuing from the left hand side of the pattern, the bases “AT” occurs twice, whereas “ATG” is unique, and therefore represents the minimum unique substring that can be found. Subsequently, it can be determined that the maximum unique substring between the pattern and the sequence is “ATGC” (Table 39).

Sequence:	ATCAAGTATGCTTAGC	
Search 1	**A**TC**AA**GT**A**TGCTT**A**GC	Five occurrences of adenine
Search 2	**AT**CAAGT**AT**GCTTA**G**C	Two occurrences of “AT”
Search 3	ATCAAGT**ATG**CTTAGC	Miniumum unique substring is identified as “ATG”
Search 4	ATCAAGT**ATGC**TTAGC	Maximum unique substring is identified as “ATGC”

Table 39: An example of the alignment algorithm used by Pindel to detect indels. Based on identifying the pattern “ATGC” in the given sequence, a stepwise pattern matching process occurs, which leads to the identification of a maximum and minimum unique substring, shared between the two sequences.

By identifying maximum and minimum unique substrings from either end of the unmapped read, Pindel then determines whether the entire mapped read can be represented by splitting the read. If the entire read can be represented in this format (i.e. an exact match, with no gaps), the putative mapping is saved. Putative read mappings are then compared, a deletion having to be supported by at least two reads. As an unmapped read containing an insertion cannot by its very nature be completely mapped, here the extra bases are recorded as the inserted sequence. Similarly, putative mappings are recorded, and must be present at least twice for an insertion to be recorded. An additional constraint placed on insertion detection is that detection of minimum and maximum unique substrings from the 5’ end is limited to read-length minus 1 (the minimum length of an insertion). Following detection, for each candidate indel, the reads at that location were then remapped to that position, to determine whether the overall mapping quality was improved in the case where the indel was present. Indels that did not improve the mapping score, were subsequently removed (Harris et al., 2013a).

7.2.1 Indel Detection with LASER

LASER (Marschall, 2013) is a more recently developed indel detection program, which similarly uses a split-read alignment method, in conjunction with the

Burrows Wheeler Aligner (BWA) package (Li and Durbin, 2010). This method similarly utilises paired end reads, where each read of a read pair (R) is split into right (R) and left (L) suffixes (S) of length M.

i.e. for each read of a read pair:

R1= S_1[1,M]	*right hand suffix of length 1 to M*
L1=S_1[l-M]	*left hand suffix of length= total read length (l) minus 1*
And	*for the corresponding read pair:*
R2=S_2[1, M]	*right hand suffix of length 1 to M*
L2=S_2[l-M]	*left hand suffix of length= total read length (l) minus 1*

Each of these four suffixes are then mapped to the reference genome using BWA, allowing up to 25 alignments to be recorded for each suffix. Suffixes for which more than 25 alignments exist are treated as unknown, and a value of M is chosen that limits the possibility of this occurring. It is highly unlikely that none of the suffixes will map at all. These anchor points, formed by those suffixes of the two reads that have mapped, define a search space for any unmapped suffixes (i.e. in this case R2 Figure 70).

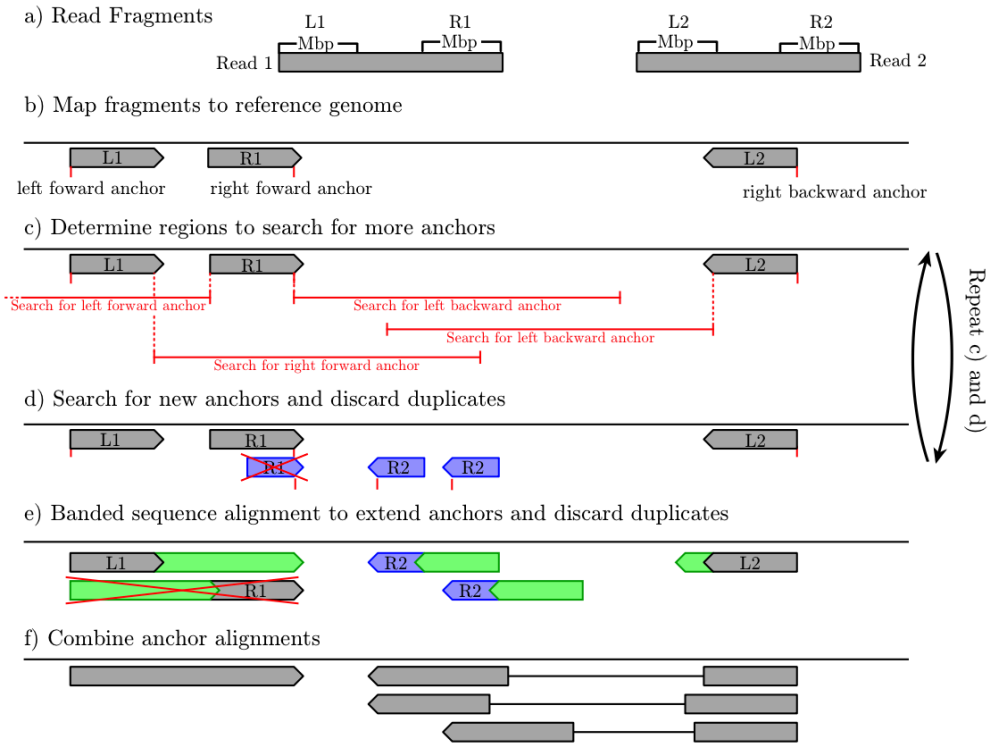


Figure 70: Schematic to aid visualisation of the indel detection process used by LASER. From the short reads (a), left and right suffixes are identified, and are mapped to the reference (b). A local anchor search is conducted three times, in order to map unmapped suffixes. Anchors are then extended and combined where reads successfully map. Otherwise, mapped suffices may be joined, artificially, representing the occurrence of an indel (f).

A local anchor search is then performed within this reduced search space defined by these anchor points. The value of M, which was originally chosen to reduce the numbers of false positive alignments being identified genome-wide is now also reduced- reflecting the assumption that the alignment of any unmapped suffixes must occur within this reduced search space. For each anchor point, a region is then searched for its corresponding suffix, and that of the mate pair. This ensures that additional anchors belonging to the same read in addition to its mate pair are identified. Additional anchor searching is repeated three times, ignoring regions already searched, so that in the extreme case where only one anchor point was found at step one, there are sufficient opportunities to identify the addition three anchor points.

Where possible, each anchor point is now extended up to the length of the complete read. When this occurs for overlapping anchor points (e.g. L1 and R1) they are combined (the read is mapped). In cases where the anchor points

cannot be extended to the length of the complete read, each partial alignments are joined by the inclusion of an indel; in this case R2 variants are joined to the L2 anchor point, indicating a deletions. Alignments are then scored based on any structural variants included or not.

To avail of the improved indel detection method utilised by Laser, and as a way to cross-reference those indels identified under the method developed previously at the Sanger Institute (Harris et al., 2013a), both methods were run in parallel, and the results compared.
The mapping results from both programs were visualised and compared in Artemis (Rutherford et al., 2000). From this, a cut-off score for indels detected by Laser (which provides a score for each alignment) was set at >10 for inclusion into this study. The output files containing indel records detected independently by Laser and from the pipeline developed at the Sanger Institute were then combined. Additional filtering of indels was carried out during this process, so that only indels that were identified in two or more taxa, or that had been detected by both methods were included. Furthermore, indels greater than 400bp were excluded, based on alignment visualisations in Artemis. Such filtering was aimed at eliminating falsely identified indels that were likely to have occurred as a result of mapping errors.

To remove genetic variation resulting from duplication changes, or that resulted from mobile genetic elements (MGEs), these regions were removed (Figure 71). In the case of MGEs an additional annotation file was provided, whereas repeats were removed by determining whether the inserted or deleted sequence was comprised of repeated substrings, or whether it matched the adjacent sequence relative to the reference genome.

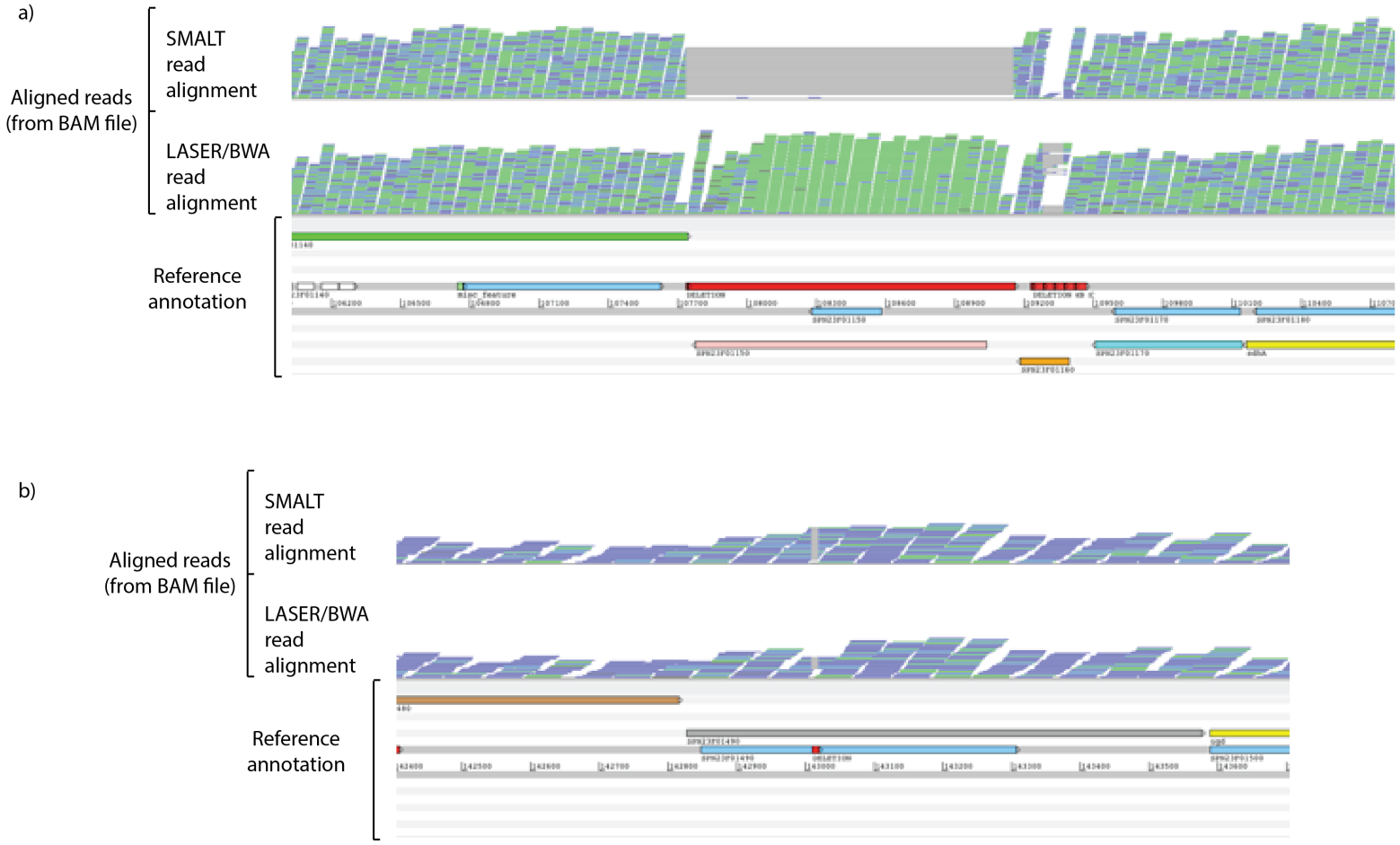


Figure 71: (a) A transposable element is indicated by the reference annotation. Mapping across this region by SMALT leads to a "deletion" being detected, due to the splitting of reads either side of this mobile element. The mapping produced by BWA in conjunction with LASER indicates reads mapped across this region. However, a much greater number of mapped reads have been coloured green in this region, relative to the normal blue, green colour ratio. This indicates that many duplicate reads occur in this region, which are concatenated and coloured green to save space in the viewer. This suggests that duplicate transposons are likely to occur elsewhere in the genome. Both cases highlight the mapping problems posed by MGEs. (b) This plots shows the occurrence of a deletion, as viewed in Artemis with the loaded BAM files. Here both SMALT and BWA/LASER support the inclusion of this deletion.

The program Gubbins was run separately to identify genome-wide recombination events (Croucher et al., 2015). The program uses a scanning statistic in conjunction with RAxML (Stamatakis, 2014) and PAML (Yang, 1997) to identify regions of abnormally high SNP density throughout the genome relative to a background rate. In addition to a file containing recombination records, a phylogenetic tree constructed in the absence of recombination events is also provided, both of which are used as input files for the software developed here.

In order to identify repair events indels were reconstructed on the phylogenetic tree provided from Gubbins. For each indel record, the common ancestor to the group of taxa possessing that indel was then identified. It was then determined whether all of the daughter isolates that diverged from that shared ancestor contained the indel. In the case where they did not, two scenarios were considered, either:

1. The same indel occurred independently, and therefore the indel was not present in the shared ancestor
2. The indel occurred in the ancestor identified, and was subsequently repaired in the downstream taxa where it was no longer present

For each branch of the tree downstream from this common ancestor the ratio of independent occurrences of the indel (scenario 1), to independent repairs (scenario 2) of the indel was determined. When an indel would need to have occurred independently more frequently than repaired to have resulted in the observed pattern of indels, a repair event was record (Figure 72).

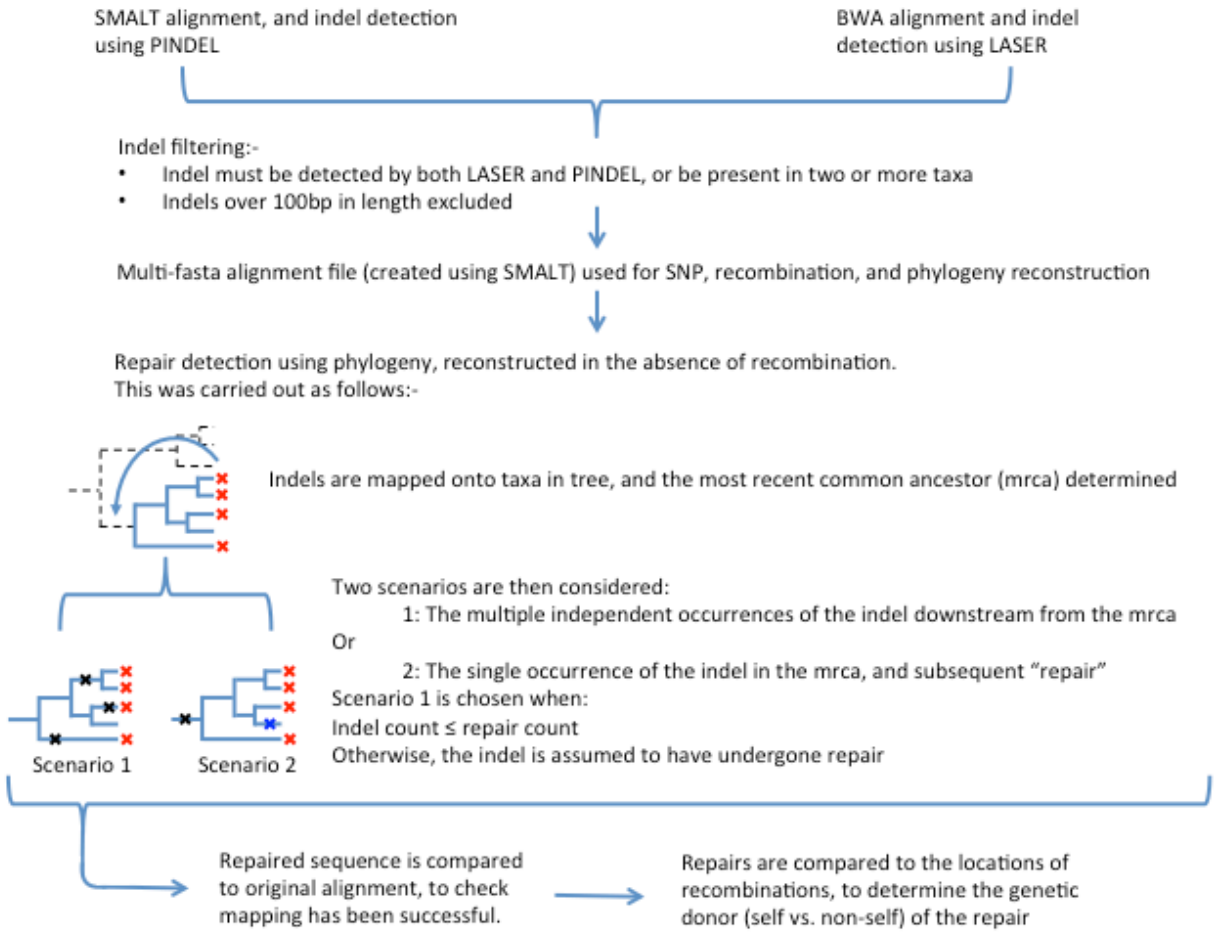


Figure 72: Schematic of the indel detection method developed here. Indels are first detected independently from the raw fastq files by the SMALT/PINDEL and BWA/LASER methods. Detected indels are concatenated, and filtered, if they are not detected by both mapping pipelines, or only occur in a single isolate. Reconstruction of the phylogeny than allows repair events to be identified, which are subject to further quality control (comparison with alignment file), before they are determined to be "self" or "non-self" events by comparison with the recombination events identified by Gubbins (Croucher et al., 2015).

Following the identification of repair events, these were then re-allocated as self- or non-self events (Figure 72, Figure 73).

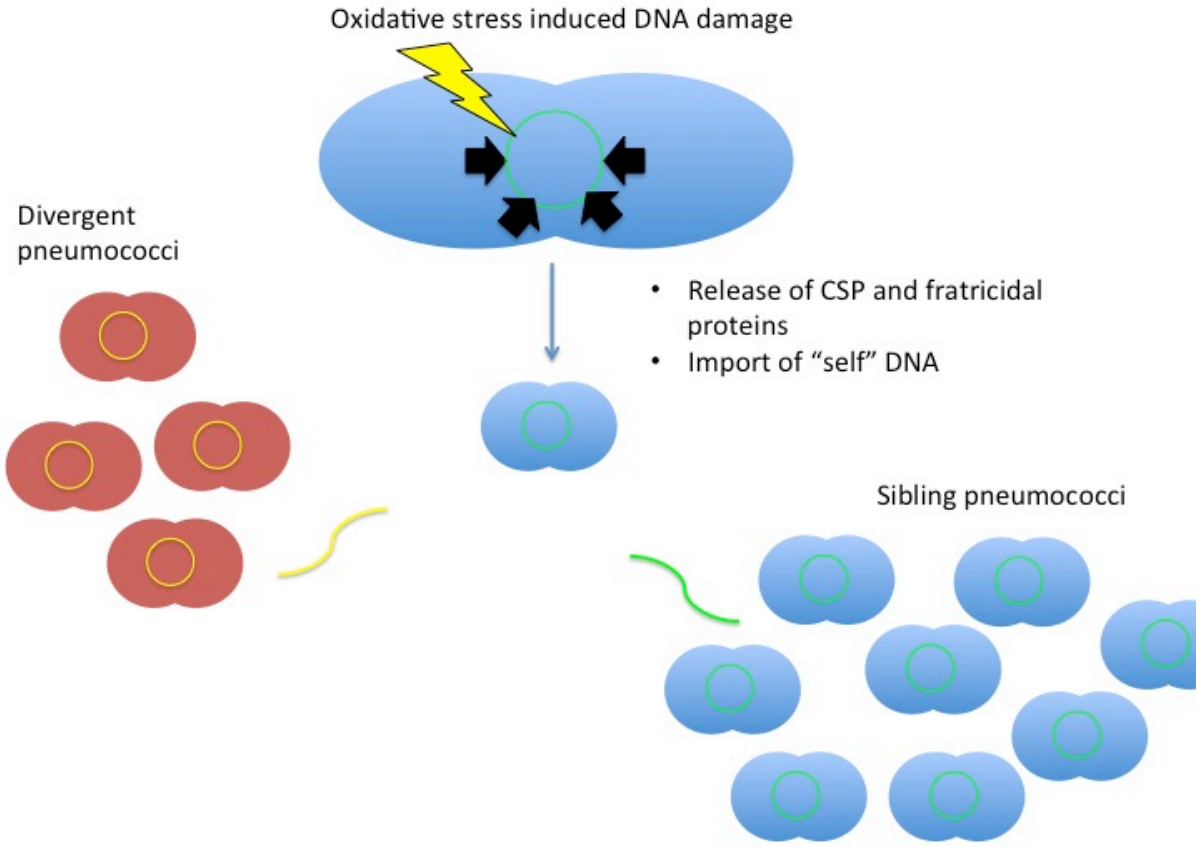


Figure 73: Self versus non-self recombination explained. Non-self occurs when a recombination event occurs between divergent taxa (bacteria coloured red), whereas self-recombination occurs when the pneumococcus recombines with other clonal members of its lineage, i.e. siblings (blue).

Reallocation of repairs was based on whether the event was found to occur within a recombination event, and using the phylogenetic reconstruction, whether both events could have occurred at the same time. In cases where this was found to be true, the event was allocated as a non-self repair, otherwise, the repair was identified as a self-repair event (Figure 73). This ratio of self to non-self was then used to reallocate insertion events on completion of repair detection. In this way it was possible to provide a conservative estimate as to the occurrence of self- versus non-self repair.

As a final output, the software developed here provided a file containing all indel, and recombination events. This was then used to manually confirm that in cases where repair events were identified, reads had successfully mapped across these regions. This was necessary in order to remove false repair events, which may have occurred due to mapping errors.

7.2.2 Datasets

Nine datasets were included in this analysis, and were diverse in terms of the sampling densities, time frames, and spatial occurrences as described below (Table 40).

The PMEN lineages (sections 1.3.8, and 2.2.5) are globally important, being frequently associated with multidrug resistance and/or a high incidence of disease (McGee et al., 2001). Four PMEN lineages were represented in this analysis. From Malawi 61 isolates belonging to PMEN27 (ST217) and isolated between 2004 and 2009 were included (Everett et al., 2012). A set of PMEN1 isolates, collected globally between 1978 to 2007 and comprised 241 sequences was also included (Croucher et al., 2011). The PMEN2 (ST90) lineage was represented by 118 globally sampled isolates collected between 1988 and 2009 (Croucher et al., 2014b). In addition the PMEN14 lineage was represented by two datasets. One was based on 53 globally sampled PMEN14 isolates (ST236) collected between 1997 and 2008 (Croucher et al., 2014a), whereas the second set comprised 70 isolates collected between 2007 and 2010 from the Maela refugee camp, located on the Thailand-Myanmar border (Chewapreecha et al., 2014a).

A further four datasets based on sequenced isolates from the Maela refugee camp were similarly included. All Maela samples were isolated as part of a longitudinal study undertaken between 2008 and 2010 on a cohort of approximately 1000 infants, and about a quarter of their mothers (Turner et al., 2012). Dense sequencing of over 3000 randomly selected isolates from this collection was carried out, resulting in a collection of roughly 100 isolates from each consecutive month of the study period (Chewapreecha et al., 2014a). Consequently, this sampling strategy differed substantially from the other datasets, which were either sampled over a longer time period, or from disparate pneumococcal populations globally. Clustering of these sequences using BAPs (Bayesian Analysis of Population structure) (Corander et al., 2003, Corander and

Tang, 2007) identified clonal clusters based on the MLST scheme, four of which were used in this study (Table 40)(Chewapreecha et al., 2014a).

Name	Origin	Sample size	MLST/ Sequence Type	Common Serotype	Other names	Antibiotic Susceptibility	Sampling period
PMEN1	Global	241	81	23F	Spain^{23F}-1	Multidrug resistant	1984-2009
PMEN2	Global	81	90	6B	Spain^{6B}-2	Multidrug resistant	1987-2009
PMEN14	Global	53	236	19F	Taiwan19F-14	Multidrug resistant	1997-2009
14 ST63	Maela	66	63	14	N/A	Low	2008-2010
19F ST236	Maela	70	236	19F	Taiwan19F-14 PMEN14	Multidrug resistant	2008-2010
19F ST4414	Maela	245	4414	19F	N/A	Low	2008-2010
23F ST802	Maela	128	802	23F	N/A	Low	2008-2010
23F ST4413	Maela	83	4413	23F	N/A	Low	2008-2010
1 ST217	Malawi	61	217	1	PMEN27	Low	2004-2009

Table 40 Datasets used in this study. For each dataset the common name if given, whether the sequences were collected globally or from a single location, whether the lineage is associated with multidrug resistance or not, and the timeframe over which samples were collected.

De novo references, used for the assembly of 23F ST802 and 23F ST4413 datasets, were constructed using Velvet and aligned using ABACAs. Annotations were transferred to these using RATT. In all other cases a reference was chosen based on publically available databases (Table 41). All sequence datasets were additionally checked to ensure that they belonged to the same ST. Where reference divergent was deemed a problem (large numbers of indels and recombination events occurring uniquely within the reference sequence), these events were removed from subsequent analysis. References used, the average mapped read length and the percentage of reads mapped are indicated in Table 41.

Dataset	Reference	Average read length	Mapping coverage
PMEN1	ATCC700669 (publicly available)	54	93.90 (2dp)
PMEN2	670-6B (publicly available)	54	95.52 (2dp)
PMEN14	Taiwan19F-14 (publicly available)	75	95.20 (2dp)
14 ST63	G54 (publicly available)	75	96.65 (2dp)
19F ST236	Taiwan19F-14 (publicly available)	75	96.63 (2dp)
19F ST4414	Taiwan19F-14 (publicly available)	75	97.07 (2dp)
23F ST802	Own reference	75	97.61 (2dp)
23F ST4413	Own reference	75	96.91 (2dp)
1 ST217	SPN1041 (publicly available)	54	96.06 (2dp)

Table 41: Datasets used in this study and the reference sequence used to during genome assembly. Average read length of samples is also indicated, which will limit the upper limit of insertion detection. As samples were sequenced at different time frames, two different read lengths occurred- primarily a function of improved technology. In addition the percentage of mapped reads is indicated.

7.3 Results

A summary of the events identified in each dataset is presented in Table 42. Three types of genomic repairs were detected: insertions, repairs, and deletions. From this a value for self to non-self repair was estimated as described above.

	PMEN1	PMEN2	PMEN14	14 ST63	19F ST236	19F ST4414	23F ST802	23F ST4413	1 ST217
Recombination	382	140	170	105	76	80	46	13	16
Recombination repairs	3	1	1	2	2	0	3	1	1
Non-self repairs	10	3	3	4	3	0	5	2	3
Insertions	49	48	39	62	37	49	91	53	34
Repairs	19	24	17	58	78	64	110	89	20
Self repairs	61	70	54	118	114	113	199	141	52
Deletions	134	95	96	105	118	161	114	108	47

Table 42: Summary of indels, recombination events and repair features identified in each dataset. Non-self repairs includes recombination repairs, in addition to insertions identified as non-self. Self repairs indicate repair events plus insertions identified as self.

Table 42 indicates most strikingly that non-self repair was identified at an almost negligible level when compared to the occurrence of self-repair. This suggest that recombination with self, is perhaps the dominant mechanism of

recombination in this species. This could furthermore indicate that pneumococci are capable of far greater rates of recombination than are currently estimated, based on SNP analysis.

It was also apparent that Gubbins identified recombination events far more frequently among the globally sampled datasets (PMEN1, PMEN2, and PMEN14), compared to those datasets collected from a single source. This could indicate that the global spread of pneumococci leads to a greater occurrence of recombination with more divergent strains. However, the sampling period was highly variable between the datasets used: the PMEN1 dataset containing samples collected over 26 years, compared to the Maela samples, which were collected between 2008 and 2010.

The effect of sampling period on the detection of these events was investigated further by dividing the number of events by the number of years sampled (Figure 74). This lead to a substantial normalisation of the datasets in terms of events identified, with recombinations appearing slightly under-represented among the globally collated datasets when adjusting for sampling period. This suggested that the initial disparity in recombination detection was largely a result of the sampling strategy, rather than a facet of PMEN biology. This is further verified by the differences in detection of events between PMEN14 and the 19F ST236 lineage of this group sampled from within the Maela camp.

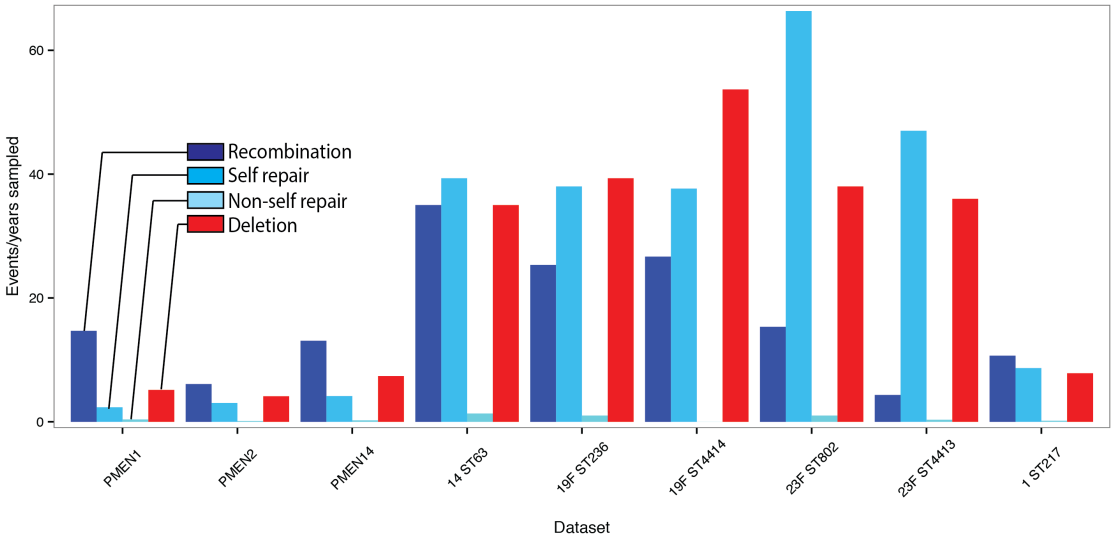


Figure 74: Plot showing the number of events divided by number of years over which samples were collected for each dataset studied.

Figure 74 also appeared to indicate that sampling over a shorter period lead to better indel discovery. This could result from both biological and methodological attributes. The dense sampling strategy used during the collection of samples from the Maela camp is likely to have lead to better lineage divergence being represented in these datasets compared to when isolates were collected over a longer time period, or from disparate pneumococcal populations. Consequently, this should favour the identification of indel repair events based on the method developed here, which relies on identifying the common ancestor of an indel and the subsequent divergence of taxa from that point. As such, in more consistently sampled datasets, such lineage divergence should have been better represented. In addition, the occurrence of the same indel across multiple taxa was used in order to improve the reliability of indel detection, similarly favouring indel detection among the more densely sampled datasets.

However, the values given in Table 42 suggest that indels were in fact identified at relatively consistent rates among all datasets used. The difference in sampling period therefore appears to have lead to an artificial increase in indel discovery in Figure 74.

The affect of sampling density on the identification of indels and recombinations was investigated further. Figure 75 shows the affect of randomly sampling taxa

from within each dataset on the cumulative number of recombination events being identified within that dataset. Sampling was repeated ten times for reliability. Here it can be seen that the relationship between recombination discovery and the number of taxa sampled is not linear. The initial steep rate of recombination detection instead begins to plateau with the inclusion of additional taxa. For those samples collected from a single source, this plateauing appears to occur at a lower level than when samples were collected globally. Interestingly, 1 ST217 is also found to plateau at a low level, despite being sampled over a longer time span than the Thai isolates.

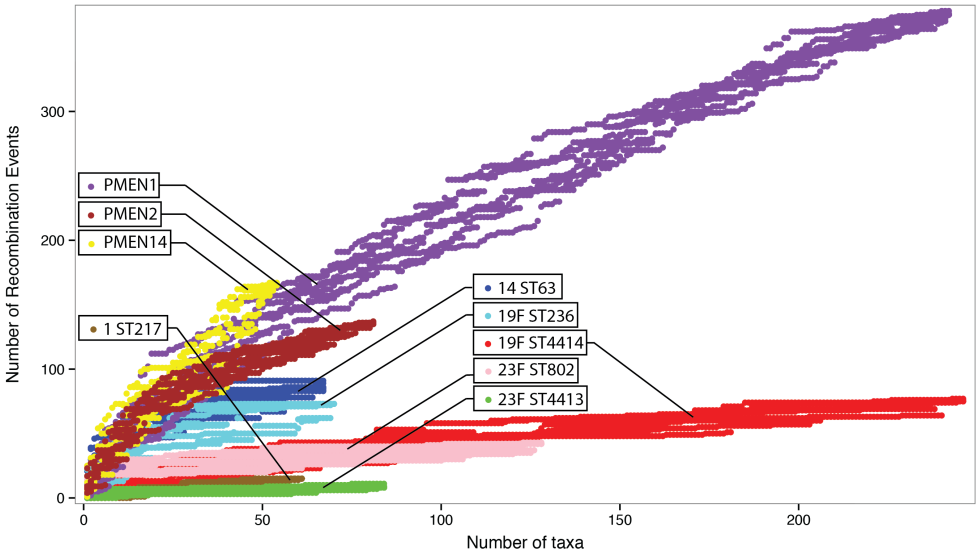


Figure 75: Simulations of recombination detection with increasing number of taxa for each dataset.

This suggests that when pneumococci are sampled globally, a greater number of recombination events are identified, but that detection of these events is to a lesser extent dependent on the length of time over which samples are collected. As such, detection of recombination appears to depend on sampling different sub-populations of pneumococci. This is supported by three factors. Firstly, when pneumococci are sampled from a single location, as in the Thai and Malawian datasets, far less recombination is identified. Secondly, the PMEN14 lineage sampled globally can be seen to follow a different trend to that which occurs when it was sampled from the Maela camp alone (19F ST236). Thirdly, although sample sizes were substantially smaller for PMEN2 and PMEN14 datasets, additional recombination events continue to be detected at a steeper

rate when compared to the locally sampled datasets. Therefore, the differences in recombination detection appear to result from differences in sampling strategy rather than having any underlying biological cause.

To determine how the identification of indels compared to the identification of recombination events, the above analysis was repeated for deletions and self-repairs. Figure 76 shows the results of these analyses, with a curve fitted, rather than presenting the raw data, in order to aid comparison between datasets (see appendix 10.11 for raw plots).

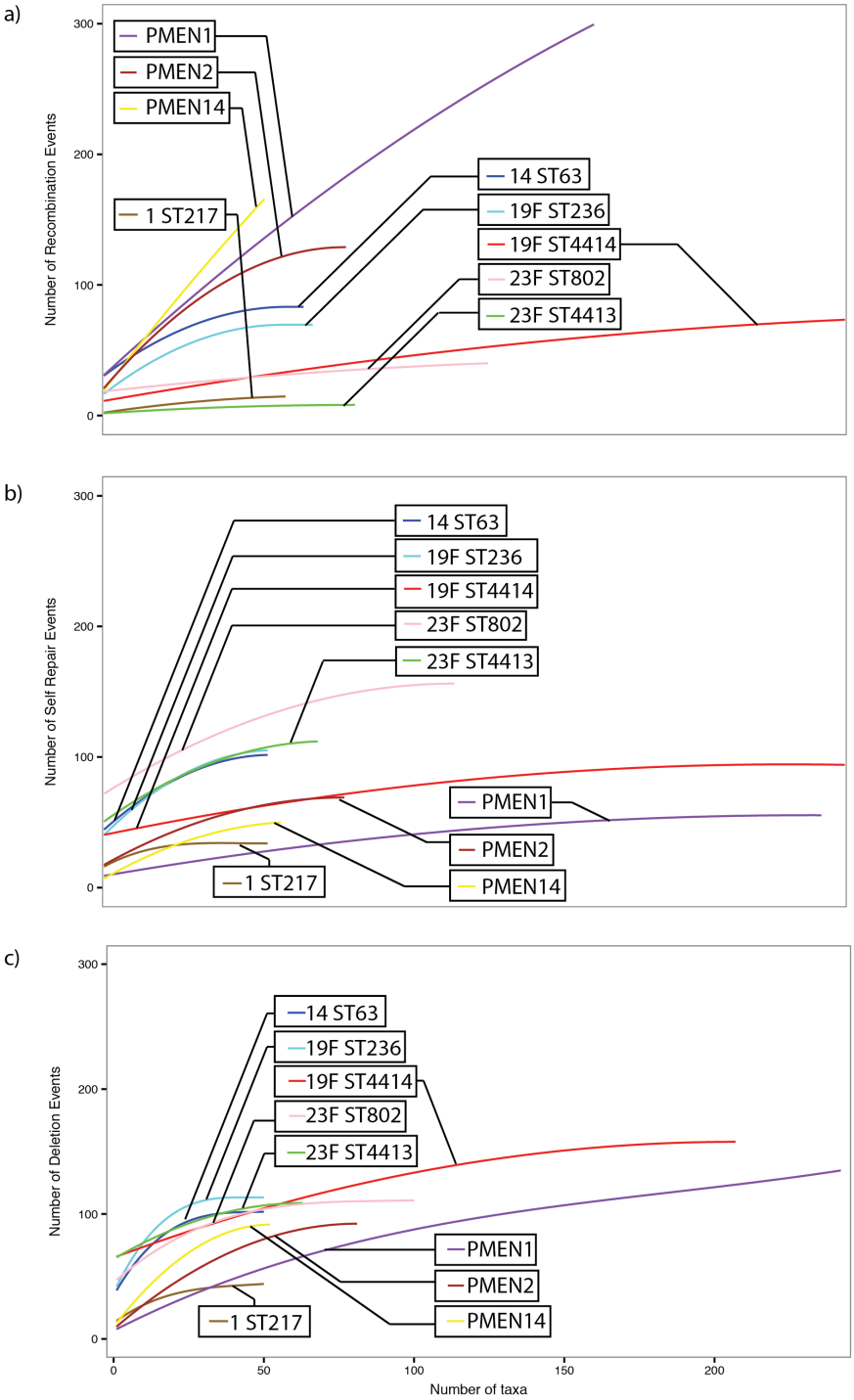


Figure 76: Simulations of recombination (a), self-repair (b) and deletion (c) detection with increasing number of taxa for each dataset dataset. Fitted curves are fitted to aid comparison, raw plots are included in the appendix (10.11).

Compared to recombination detection, the identification of deletions and repairs plateaued at a more consistent level for all datasets (Figure 76). In addition, initial indel discovery was often far more rapid, than the identification of recombinations. This was particularly evident when comparing the datasets

collected over a longer time period (1 ST217, PMEN1, PMEN2 and PMEN14), compared to the Maela camp datasets. Here, these form two distinct groups. This is evident from the starting point of the fitted curve, which occurs at a higher level in those curves fitted to the Maela camp datasets. This indicates that indels, to a greater extend than recombinations, tended to be conserved across most members of the dataset. This also explains the plateauing of the curve as more taxa were sampled. For datasets collected over a longer time period, the fitted curve was shallower, and could be fitted closer to the origin. This suggested a greater detection of independent indels occurred in these datasets. In summary, sampling density appeared to influence the discovery of indels, in contrast to detection of recombination, which was more influenced by sampling widely. This indicated that when isolates were densely sampled, i.e. over limited time periods, a greater number of indels could be detected. If the occurrence of indels predominantly led to fitness costs in the host, this trend might be expected. Under this scenario, isolates that accumulated indels would likely be purged from the population over increased time periods. Whilst deletions would be expected to be a predominantly deleterious force, the number of repair events followed a similar trend.

To investigate the potential fitness effects of indels occurring in the genome, for each dataset, deletions and repairs were compared to the reference annotation to determine whether they occurred within a CDS. Based on this analysis, of the 976 deletions identified, 45% were found to have occurred within CDS regions. In addition, 66% of these events were predicted to have resulted in a frameshift (Table 43). In order to test the affect of time period over which samples were collected on the detection of these events, deletions and repairs occurring in PMEN1, PMEN2 and PMEN14 datasets were compared to those sampled exclusively over a limited period from within the Maela refugee camp (Table 43). Whilst this indicated that the occurrence of deleterious indels was lower among isolates sampled over a longer time span this affect was not strong. This in part could result from gene paralogs being present elsewhere in the genome. Consequently, some frameshifts may have been compensated for more effectively. In addition, no *in vitro* verification was carried out in this analysis to

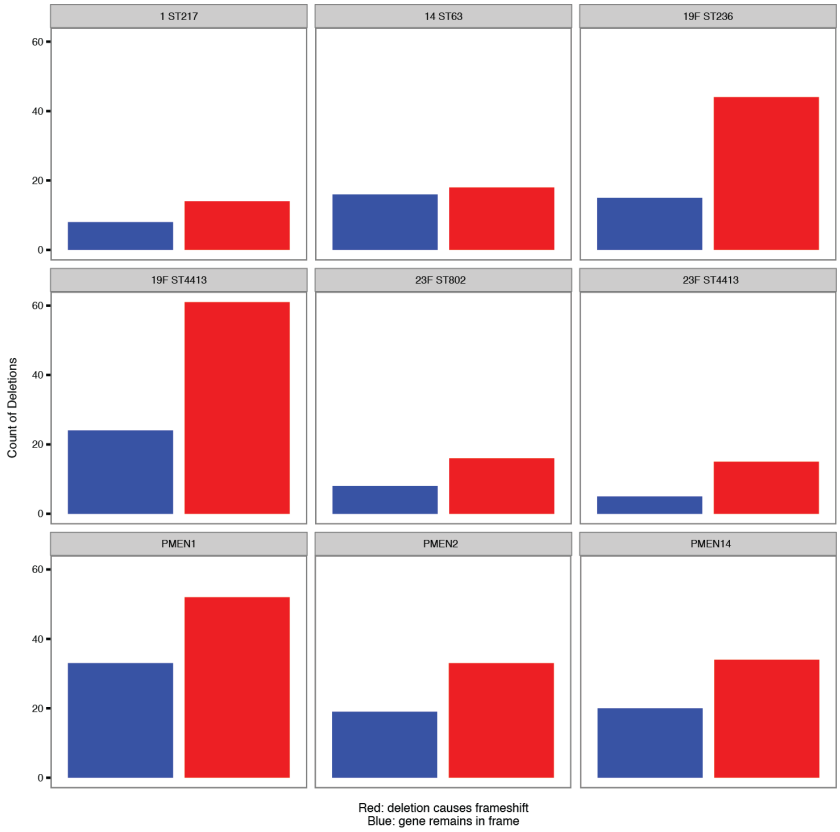
confirm whether such events truly inactivated the genes in which they occurred. This is perhaps why a stronger trend was not identified.

Dataset	Count of deletions	Total number of deletions in genes	% of deletions in genes	Count of those causing frameshifts	% causing frameshifts
Maela camp	608	222	37	154	69
PMEN global sets	322	191	59	119	62
	Count of insertions	**Total number of insertions in genes**	**% of insertions in genes**	**Count of thise causing frameshifts**	**% causing frameshifts**
Maela camp	649	176	27	124	70
PMEN global sets	179	72	40	46	64

Table 43: The occurrence of indels within genes based on the reference annotation, for isolates sampled over a longer time period (PMEN global sets) compared to those isolates densely sampled over a short time period (Maela camp).

To determine the consistency of this affect, the occurrence of indels within genes, and whether these were expected to have resulted in a frameshift or not was plotted for each datasets (Figure 77).

a)



b)

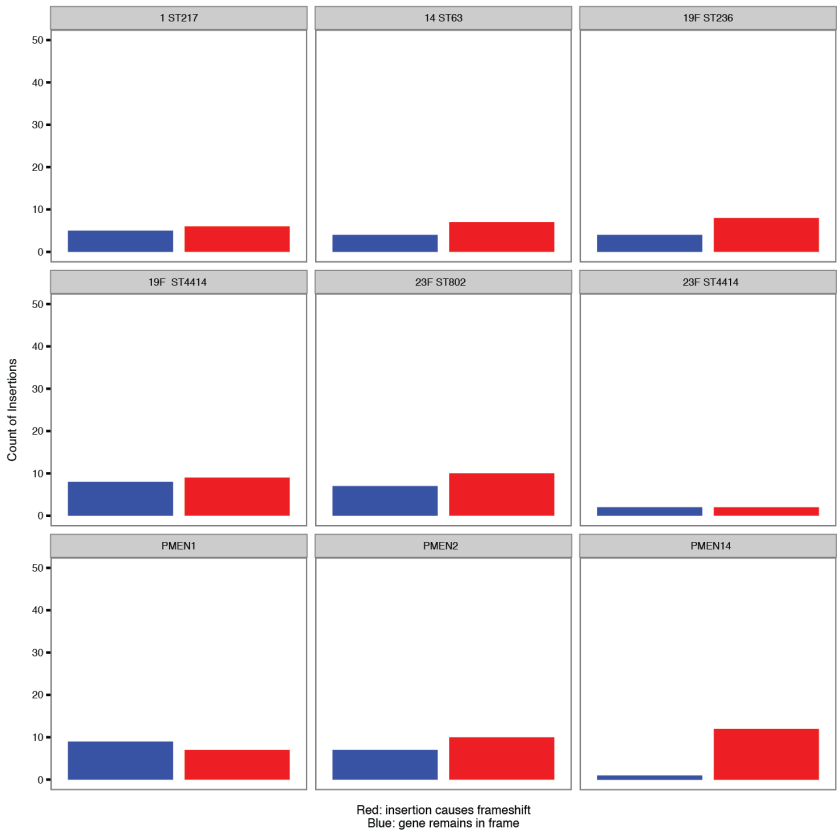


Figure 77: Raw counts of the occurrence of deletions (a) and insertions (b) in genes, and the frequency with which these were calculated to have resulted in a frameshift within a functional gene (red= frameshift predicted, blue= gene remains in frame).

Whilst little consistency was found between datasets, far fewer insertions appeared to occur within CDS regions compared to deletions. This could indicate that repair events remained under-detected in this analysis. The greater rate of deletion recorded within genes would be expected to lead to rapid genome degradation and loss of fitness if this was not counteracted by the repair of such damage. Therefore, this is supportive of the repair detection method developed here being conservative.

In most cases insertions, like deletions, were predicted to result in a frameshifts when they occurred within a CDS. This was particularly pronounced in the PMEN14 dataset (Figure 77). Whilst this might seem slightly paradoxical to the hypothesis that these are repair events, there are a number of reason why this might be predicted. Under the hypothesis that insertions represent repairs, when such events occur within a functional gene, it would imply additional modification has occurred elsewhere in that gene. Consequently the "repair" actually results in a frameshift. This could also reflect methodological inaccuracies in the ability to detect the precise length of an insertion event, which unlike deletions is reliant on aligning a shorter proportion of the read length. Consequently, whilst insertions appear to predominantly lead to CDS disruption, the ability to investigate this further is once again limited by the read length currently offered by NGS technology.

7.4 Deletions in detail

As discussed previously, deletions in *E. coli* are marked by a Chi sequence (Kulkarni and Julin, 2004). An analysis was therefore undertaken to determine whether similar sites could be detected in pneumococci. Given the *E. coli* Chi site is eight nucleotides long, a slightly longer sequence of ten bases up and downstream of each deletion event was subject to investigation. Firstly a count of the bases present was made to determine whether any particular nucleotide was particularly prevalent in this region, and whether the GC content of this sequence differed from the expected ~39% GC content commonly recorded for pneumococci (Tettelin et al., 2001, Croucher et al., 2009). Table 44 indicates the

output of this analysis, which suggested that GC content was not substantially different from the expected value of approximately ~39%. Furthermore, no particular nucleotide appeared particularly prevalent, or under-represented in the sequence prior to or following a deletion event.

Upstream sequence				
Count of "A"	Count of "T"	Count of "G"	Count of "C"	
57710	56327	32321	34397	
Calculated AT content		Calculated GC content		% GC content
114037		66718		37%
Downstream sequence				
Count of "A"	Count of "T"	Count of "G"	Count of "C"	
55689	56327	32321	34397	
Calculated AT content		Calculated GC content		% GC content
112016		66718		37%

Table 44: Comparison of the base frequencies present in the 10 nucleotides immediately prior to and following a deletion event. Nucleotide contents are calculated across the total bp, and the GC content considered shown, the normal value across the pneumococcal genome being 39% GC content.

The upstream and downstream sequences were investigated further by determining the frequency at which each base appeared at positions prior to, and following the deletion event (Figure 78).

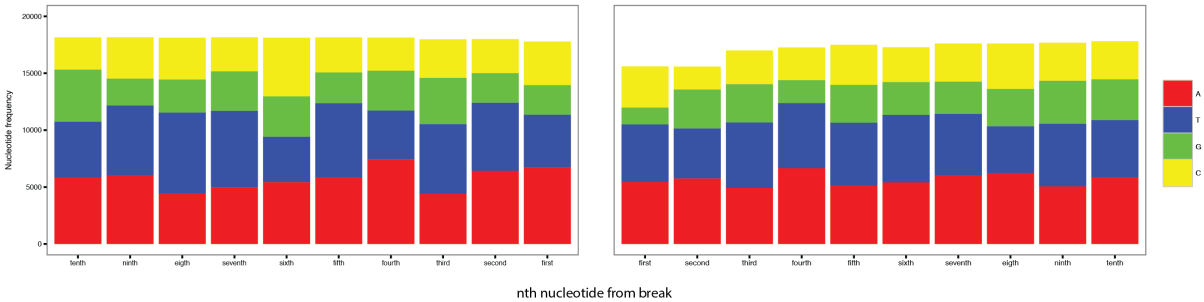


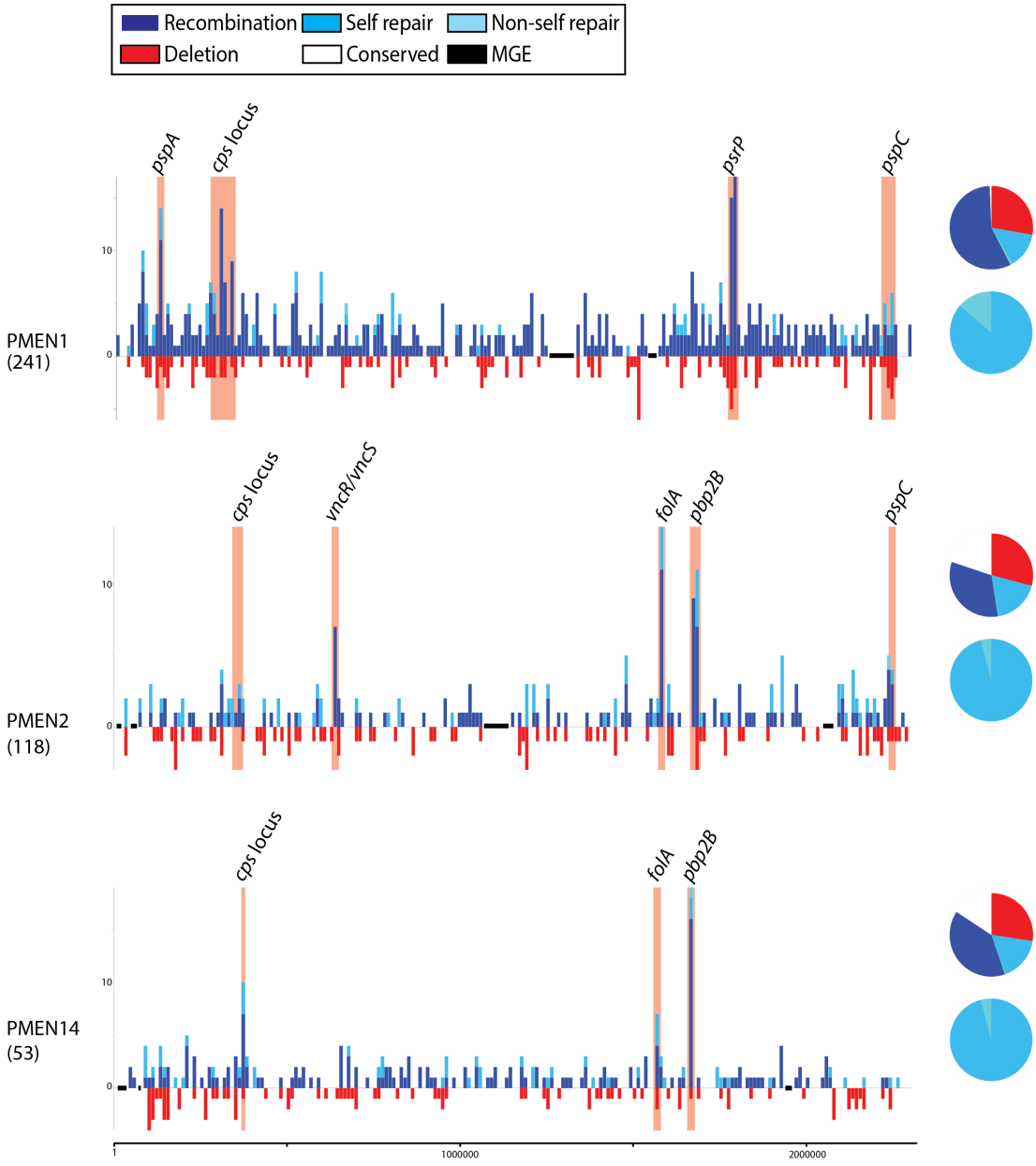
Figure 78: Base frequencies at each location leading up to and following the occurrence of a deletion event.

Consistent with Table 44, there was no obvious difference in the base frequencies in these regions, apart from in the case of guanine, and no conserved sequences could be identified visually. As such no Chi site analogy was found.

Guanine appeared to be less frequently detected in the first base following the occurrence of a deletion however. Interestingly guanine possesses the lowest oxidation potential of all the DNA bases, and consequently is particularly susceptible to oxidation (Evans and Cooke, 2004). As such, the low frequency of this base immediately following a deletion event could be taken as an indication of oxidative damage.

7.5 Genome-wide analysis

To determine how the distribution of deletion, repair and recombination events differed with distance from the replication origin, plots were constructed by dividing the genome into sections of 10,000bps, and making counts of the numbers of each event occurring within these sections (Figure 79). Consistent with previous analyses (Croucher et al., 2011, Croucher et al., 2014a, Croucher et al., 2014b, Chewapreecha et al., 2014a) several recombinational hotspots were identified. This was particularly evident for the *cps* locus, due to the selective pressure for capsule switching, and beta-lactam resistance. In contrast deletion appeared to occur relatively randomly throughout the genome, and was not clustered near the origin of replication, as would be expected if this process resulted from replication errors. This relatively random distribution of genetic damage was therefore supportive of a scenario of random oxidative damage occurring throughout the genome.


Recombination
Self repair
Non-self repair
Deletion
Conserved
MGE
pspA
cps locus
psrP
pspC
PMEN1
(241)
cps locus
vncR/vncS
folA
pbp2B
pspC
PMEN2
(118)
cps locus
folA
pbp2B
PMEN14
(53)
10
0
1
1000000
2000000

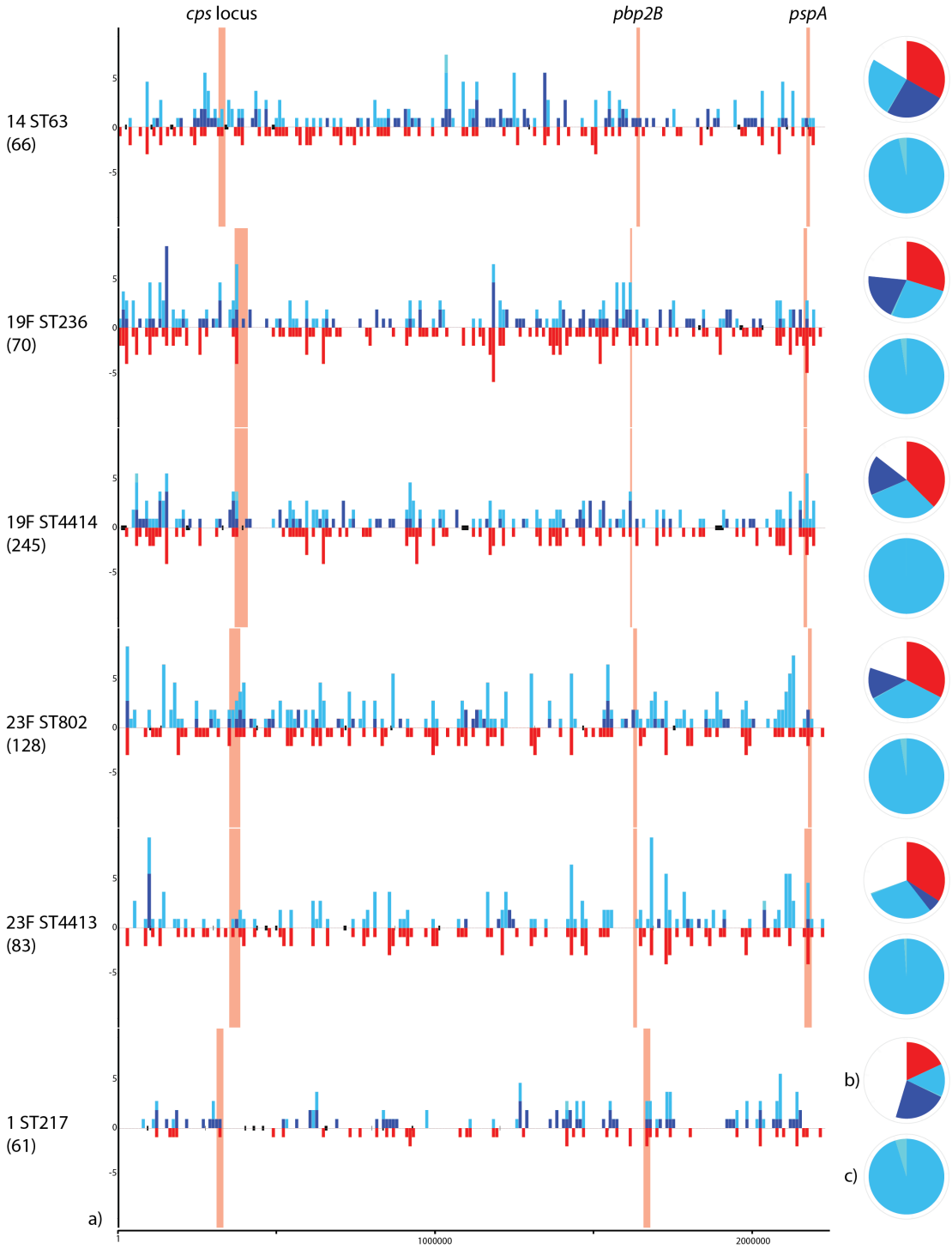


Figure 79: The occurrence of deletion (red), recombination (dark blue) and repair (self= blue, non-self= cyan). For each dataset, a heatmap was plotted to indicate the occurrence of these events with distance from the replication origin, (b) a pie chart showing the proportion of the genome undergoing such events, (c) the ratio of self to non-self repair. Positions of MGEs are also indicated, events being removed from these zones (coloured black on the heatmaps).

During this analysis, pie charts were also constructed to determine the proportion of the genome subject to these events, and the ratio to which self- versus non-self repair was detected for each dataset. Here deletion can be seen

to have affected a large proportion of the genome, with recombination and insertion events being more restricted in their occurrences.

A relatively small proportion of the genome appeared to be conserved in most datasets, with PMEN1 an extreme case. As discussed previously however, this is likely to be influenced by the sampling regime. Furthermore, it might be expected that certain environmental constraints place greater or lesser pressures on populations to maintain genomic integrity depending on the environment.

To determine the occurrence of these events at the taxonomic level, for each taxa, the number of recombinations, repairs and deletions was plotted, alongside their position in the phylogeny, which was constructed in the absence of these events.

For PMEN2, this analysis was particularly informative (Figure 80). Here, deletion, repair and recombination appear to occur relatively random between all isolates, apart from a single lineage that invaded Iceland.

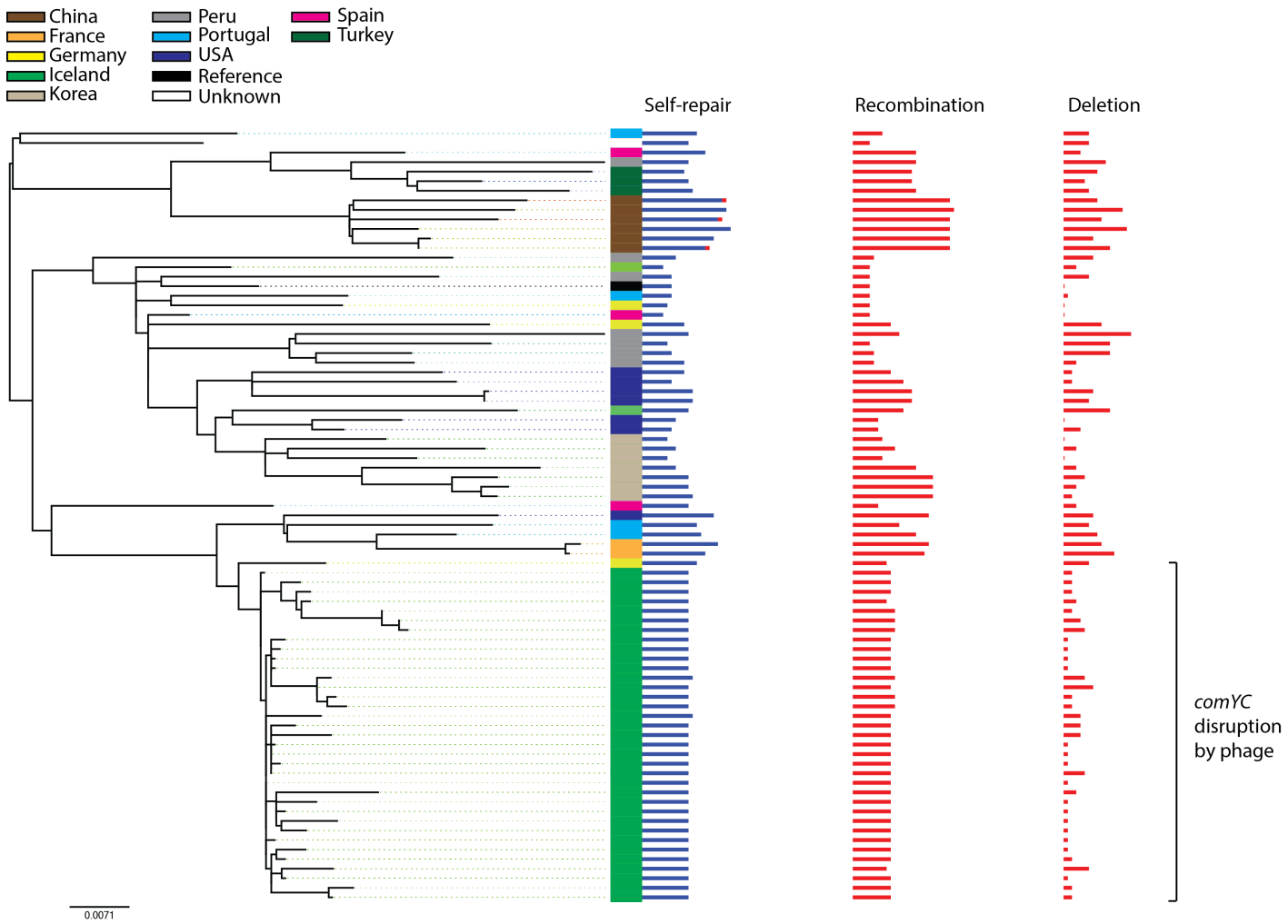


Figure 80: The demise of PMEN2 in Iceland has previously been attributed to competence disruption through phage insertion (Croucher et al., 2014b). Here self-repair can similarly be seen to have occurred at the same rate, independent of the position of an isolate in the phylogeny. Furthermore, since the introduction of this lineage into Iceland, one additional self-repair event appeared to have occurred. The subsequent demise of this lineage, despite a decline in antibiotic usage in Iceland at the time, is therefore likely due to fitness costs associated with the inability to repair genetic damage.

The lineage present in Iceland can be seen to have undergone large amounts of self-repair. However, most of the isolates within this group contained the same 11 self-repairs, with two isolates containing an additional repair event. As such, it appears that on entering Iceland, this lineage no longer underwent self-repair.

The decline of PMEN2 from Iceland, has recently been correlated with the occurrence of a phage, that resulted in the disruption of the *comYC* gene (Croucher et al., 2014b). This gene has previously been identified as part of an essential component of the competence system (Croucher et al., 2011). Croucher and colleagues (2014b) have previously suggested ComYC disruption prevented this lineage from recombining, and being able to respond to clinical pressures. This analysis suggests that a lack of self-repair coupled to a loss in competence could similarly have lead to the decline of this lineage from Iceland. Repair analysis may therefore be informative as to the likely persistence and prevalence of pneumococcal lineages.

7.6 Discussion

The study described here covers new ways to improve the accuracy of indel detection. In addition a method for the identification of repair events was developed, which avails of the high throughput sequencing technologies that have allowed isolates to be sequenced and analysed in their hundreds. Using this approach it has been shown that deletion events can be identified genome-wide, and that such genetic disruption does not appear to result from replication errors. The widespread occurrence of deletion implies that pneumococci must have a mechanism of repairing such damage. Genetic repair was identified using two approaches:- by identifying the physical insertion of genetic material into the genome, and through the identification of events in which an indel was present in the ancestor to a group of taxa, but was not present in all of the descendants of this progenitor. It was further demonstrated, using previous SNP based recombination analysis (Croucher et al., 2015), that recombination with divergent taxa (non-self repair) rarely resulted in genetic repair. Instead, self-repair appeared to be the dominant method of removing genetic damage. In the

PMEN2 lineage, where this process was prevented through disruption of the *comYC* gene by phage, it is possible that the accumulation of deletions and mutations caused this lineage to fall into decline.

7.6.1 Indel detection

The method used to detect indels in this study was based on split-read detection. As such, indel detection is expected to be conservative, as other methods of identifying these events are available. Part of the reason these additional indel detection methods were not employed here was due to the complications associated with accurate indel identification. Although genome assembly is now undertaken routinely, and is largely automated, mapping errors, and abnormal coverages still occur relatively frequently. Read length therefore remains a major limitation of current sequencing technology. For this reason, the method employed here was limited to previously tested indel detection techniques (Harris et al., 2013a). In addition, indel detection was limited to a maximum length of 400bp, with repeat regions, and mobile elements removed- in order to further reduce the chance of false indel detection. Consequently, it is expected that the figures quoted here are an underestimate of the true level of recombination and genetic damage occurring among pneumococci *in situ.*

A further complication in the detection of both SNP based recombination and indels is that detection will tend to saturate as a greater proportion of the genome is found to undergo such events. The problem of recombination within previous recombinations and indels alike, will therefore limit the ability to continue to detect these process at higher levels. The analysis presented here on the affect of sampling strategy on the detection of these events is therefore particularly informative, and should be considered prior to sequencing isolates depending on the questions being asked i.e. whether indel detection, or recombination is of primary interest.

7.6.2 Causes of Deletions

It has been discussed how genetic deletion can arise from different sources, particularly in the context of replication errors and due to oxidative stress. The apparent random distribution of deletion events throughout the genome therefore suggests that the deletion events identified within this study are most likely attributed to oxidative stress (Pericone et al., 2002). If such events occurred due to replication errors, they would be expected to occur predominantly close to the origin of replication, a trend which was not found. Although small, the reduction in guanine residues immediately following the occurrence of deletion events identified in this study would also support this finding, due to its high susceptibility to oxidative damage (Evans and Cooke, 2004).

In addition pneumococci are known to be subject to high levels of oxidative stress during nasopharyngeal colonisation. Hydrogen peroxide is produced as part of pneumococci's metabolism, and it is exposed to many additional sources of free radicals within this niche (Pericone et al., 2002, Kohanski et al., 2007). In chapter 6, despite exposure to hydrogen peroxide however, very little genetic damage was identified. This suggests that the method of administrating this stress was not representative of the natural state. This is likely because of the high reactivity of the hydroxyl free radical that must be generated close to genetic material in order to react (Redmond and Kochevar, 2006). For this reason, it will be important to test the level of deletion and repair that occurs in pneumococci subject to differing levels of oxidative stress in natural conditions. Under conditions of elevated stress, it would therefore be predicted that to counteract the increase in genetic damage manifested through deletion, competence would need to be up-regulated. Sickle cell disease patients may offer an important opportunity to study this process, owing the high levels of oxidative stress that occur internally as a result of this condition (Carter et al., 2014).

7.6.3 Chi sites and End joining machinery

Despite the occurrence of Chi sites in some bacteria (Kulkarni and Julin, 2004), no evidence for such a sequence could be identified among the pneumococci datasets used in this study. The pneumococcal genome is commonly regarded as having a low GC content, and the ~39% GC content commonly cited for this species (Tettelin et al., 2001, Croucher et al., 2009) appeared to be present up and downstream from the site of a deletion. The only variation of note, was a low guanine frequency in the base immediately adjacent to the break point, located in the upstream direction, as discussed previously. As such no Chi-like sequence could be identified based on nucleotide frequencies adjacent to the deletion, or by manually viewing these regions. An enzyme for non-homologous end-joining also remains to be identified for the pneumococcus, although the deletion analysis carried out here indicates that one does exist.

7.6.4 Indels as repairs

As expected indels predominantly occurred in intergenic regions of the genome, where presumably they impose less of a fitness cost on the host organism, and consequently are better tolerated. Despite this, both deletions and genetic insertions of sequences appeared to predominantly result in frameshift mutations. This implies that although insertion should represent a repair mechanism, the insertion of such material is not always beneficial to the host. This finding currently remains paradoxical, and is either a reflection of prior genetic modification within the CDS region, which therefore effectively invalidates any subsequent repair, or is related to current limitations on the accurate identification of indels. As such there is a need for longer read lengths to better decipher between biological and technological artefacts.

7.6.5 Sample density

Both indel and recombination detection was found to vary according to the sample strategy employed. Densely sampling pneumococci from a single location over a short time period was found to improve the ability to detect indels, whereas more recombination was identified when sampling occurred widely

between different pneumococcal lineages. Interestingly, the relationship between sample size and the detection of these events was not linear. Instead, the effect of increasing sample size on the detection of such events decreased as more taxa were sampled. This suggests that in many cases, indels and recombinations were shared among isolates.

7.7 Conclusions

This analysis indicates self-repair is an important mechanism by which pneumococci can repair deletion damage, likely caused as a result of oxidative stress whilst inhabiting its niche. The identification of such events can be extended to the taxonomic level. Mapping these events across the phylogeny of PMEN2 indicates, that where this recombination is prevented, there is the potential for genomic degradation to occur. The resultant fitness costs could help explain the decline in prevalence of this lineage in Iceland.

8 Final Discussion

The pneumococcus remains a pathogen of global concern, estimated to cause 1.6 million deaths annually, with between 0.7 and 1 million of these occurring in children below the age of 5 (WHO, 2015). Antimicrobials resistance, in conjunction with serotype switching and replacement has allowed the pneumococcus to thwart control measures, driving the need for continuous redevelopment of intervention and treatment measures. As a consequence of the high cost of controlling and treating pneumococcal disease, the burden of pneumococcal infection has become increasingly placed on developing countries that typically lack the medical resources that are available in the developed world.

The global incidence of pneumococcal disease has led to a drive to better control this pathogen in resource poor countries such as Malawi, facilitated more recently by charitable donation and global alliances such as GAVI. Whilst such support is welcome, often lacking an understanding of the local epidemiology and risk factors for disease, the short and long-term effectiveness of such intervention strategies remain an active area of applied research and hypothesis testing. Of particular interest is the effectiveness of intervention in areas of high HIV incidence, and with different serotype distributions and population dynamics than are typically seen in countries such as the USA, where interventions, such as the PCVs, were first developed and marketed.

In Malawi, such control measures have included the introduction of ART, and PCV13, while ceftriaxone also became freely available to the country from 2003. Active disease surveillance has made a valuable contribution to predicting the long-term effectiveness of such measures. In addition clinical trials have formed an important component of understanding the effectiveness of vaccination in HIV sufferers (French et al., 2010), and the cost-effectiveness of such interventions (Costello, 2015). More recently, genetic analyses have begun to be undertaken, with a view to assessing the sequence and serotype dynamics in

circulation, as well as the mechanistic basis of antimicrobial resistances in Malawi.

Such analyses predict PCV13 to protect against approximately 60% of invasive pneumococci currently in circulation. However, the potential for serotype replacement in the future was also indicated (Everett et al., 2012). Of further concern is the spread of penicillin resistance (Everett et al., 2011) and the recent emergence of ceftriaxone resistance (pers. comm. Everett 2015). The occurrence of globally circulating pneumococcal lineages in Malawi was similarly demonstrated, indicating the potential for MDR resistant lineages to contribute to resistance in Malawi (Everett et al., 2012). However, the mechanisms by which beta-lactam resistance arises clinically remain poorly understood. Transformation with genetic material acquired from co-colonising oral streptococci appears to play a key role in this process. However, the frequency of such exchanges, and the potential for viridans streptococci to donate resistance associated genetic material to pneumococci has not been assessed.

Despite the importance of recombination in promoting the genomic plasticity of the pneumococcus it is only relatively recently, facilitated by the increasing availability of NGS technologies, that this process has begun to be understood at a population level. The analysis of globally important MDR pneumococcal lineages, has implicated the importance of recombination in the ability for pneumococci to respond to clinical intervention (Croucher et al., 2011, Croucher et al., 2014a). In addition, the effect of disruption to recombination machineries, as a result of phage insertion, has been found to affect the long-term viability of pneumococcal populations (Croucher et al., 2014b). Cross-lineage analyses have further suggested different rates of recombination occur between pneumococcal clones, which could affect the potential for such clones to acquire MDR (Hanage et al., 2009). Whilst resistance may be driven by recombination, resistance invoked *in vitro*, frequently results in a fitness cost. The MDR PMEN lineages are consequently of greatest interest and threat, as they appear to have developed drug resistance, while maintaining fitness relative to susceptible pneumococci (Rudolf et al., 2011). Although recombination appears to be the driver for the

rapid development resistance in such lineages, amelioratory mechanisms such as mutation likely play an important role in compensating for the fitness costs associated with drug resistance (Orio et al., 2011).

Despite a long history of study, the mode by which beta-lactam resistance arises in the apparent absence of fitness costs remains unclear. This is partly due to the erroneous inference of resistance *in vitro* as a model for clinical resistance. Whilst there has developed an increasing appreciation of studying beta-lactam resistance among clinical isolates, analyses have typically been subject to studying the interaction between beta-lactams and target PBP proteins, rather than considering the wider genomic context as a whole (Hakenbeck et al., 2012). It is only with the availability large genome sequence databases that beta-lactam resistance has began to be investigated genome-wide and across bacterial populations (Chewapreecha et al., 2014a). In addition to understanding the genetic basis for beta-lactam resistance, the increasing availability of publicly available sequence databases offers the opportunity to assess the role that globally circulating populations of bacteria have had on the emergence of beta-lactam resistance. This includes the potential to assess the exchange of genetic material between pneumococci and viridans streptococci, in addition to other globally circulating pneumococcal populations.

Whilst recombination plays an important role in drug resistance and serotype switching, this process also performs a wider role in the repair of genetic damage (Vos, 2009). Genetic damage has typically been viewed as the occurrence of mutation, which must be purged in order for fitness to be maintained. However, the pneumococcus exists in a highly stressed environment, owing to the abundance of oxidative free radicals, derived from it's own metabolism, immune attack, and as well as artificial agents, such as antibiotics (Pericone et al., 2000). Oxidative stress, as a result of such assault, is widely documented to cause single and double strand DNA breaks, which can be deadly to the cell (White et al., 2009, Dharmadhikari et al., 2014, Yesilkaya et al., 2013). Indeed, the millimolar concentrations of hydrogen peroxide produced by the pneumococcus appears sufficient to inhibit the growth of other

nasopharyngeal co-colonisers such as *H. influenzae* (Pericone et al., 2000). Despite the widespread occurrence of oxidative free radicals, the pneumococcus appears surprisingly resilient to the genomic degradation frequently associated with such agents. Furthermore, the pneumococcus appears to lack common peroxide scavengers, and mechanisms such as the SOS response that commonly protect and repair oxidative damage (Yesilkaya et al., 2013).

Whilst recombination has become routinely documented across pneumococcal lineages, the extent to which deletion occurs and is repaired within this species has not been characterised.

The aim of this thesis was to characterise the role of recombination, as a mechanism of deletion repair, and in the spread of beta-lactam resistance, on a regional and global scale. This included the identification of recombinational exchanges occurring both intra- and inter-species, within globally important pneumococcal clones as well as mixed, single source pneumococcal populations.

8.1 Beta-lactam resistance

Beta-lactam resistance is primarily driven by the modification of *pbp* enzymes (Zapun et al., 2008). These enzymes interact directly with the beta-lactam, which mimics the *pbps* natural D-alanyl-D-alanine substrate. Malawian *pbp1a, pbp2b* and *pbp2x* genes demonstrated high levels of genetic diversity, indicative of recombination in response to antibiotic pressure (Chewapreecha et al., 2014a). Whilst beta-lactam resistance is often found to correlate with *pbp* divergence, much of the allelic variation identified here seemed to have little affect on penicillin susceptibility, which generally clustered poorly among the more divergent alleles. Identifying recombination across these genes indicated that these events were not restricted to key sites, but occurred throughout the gene. Consequently, only recombination events that introduced resistance associated SNPs into the more sensitive regions of the gene, such the conserved SxxK, SxN and KxG motifs, are likely to affect overall beta-lactam susceptibility. However, whilst only a limited amount of the recombination led to the introduction of

resistance-associated SNPs based on the literature, it was apparent that the loss of such SNPs did not necessarily correspond with a drop in penicillin susceptibility *in vitro*. This indicates the importance of the wider genetic background in the expression of such modifications.

In accordance, the GWAS method developed identified a number of genes outside of the cell wall synthesis pathway, which were subsequently screened for recombination, and sequence mosaicism. This identified the cell wall synthesis genes *ddlA* and *murF* as carrying SNPs associated with a loss in beta-lactam susceptibility. Despite this, due to their close proximity to the highly recombinogenic, and beta-lactam resistance associated *pbp2b* gene, it remains unclear the degree to which such associations are an artefact of hitchhiking across this region (Smith and Klugman, 1998). Of particular interest however was the association of *leuS* SNPs with beta-lactam non-susceptibility. Interestingly, this gene was found to contain SNPs that associated statistically with beta-lactam resistance, and contained sequence blocks sharing a high degree of similarity to *S. mitis,* suggesting a role for inter-species recombination. This analysis is the first to indicate the potential involvement for the *leuS* gene in clinical beta-lactam resistance among pneumococci.

Interspecies recombinational exchanges were also identified widely among *pbp* alleles. In total 23 unique resistance-associated SNPs, based on the literature, were found to have been introduced across the three *pbp* genes during a number of different recombination events. This indicated that exchanges between *S. mitis* and *S. oralis* appeared to occur far less frequently than between other pneumococci. Whilst this trend is likely to have been largely exaggerated because of a deficit in viridans streptococcal sequences available for study, the attempt to isolate co-colonising *S. mitis, S. oralis* and pneumococci from nasopharyngeal samples was similarly unsuccessful. Despite the inability to detect recombination between these species *in vitro,* and the limited occurrence of such exchanges detected *in silico*, genetic transmission of resistance between these species has clearly played an essential role in the acquisition of resistance by pneumococci (Dowson et al., 1993). Consequently, such exchanges, although

facilitated by the pneumococci's mismatch repair system (Humbert et al., 1995) perhaps occur infrequently, and subsequent transmission between pneumococci drives the dissemination of resistance. It is similarly possible that such exchanges occur within a particular subset of the population, which was not specifically targeted in this study. Infants (<2yrs), where co-carriage of pneumococci and these other species has been found to occur, would be an important subset for future sampling (Kononen et al., 2002).

It was apparent from the distribution of these SNPs that beta-lactam resistance was not occurring by any one, or by any particularly prevalent mechanism, suggesting that penicillin resistance is occurring by many different routes in Malawi. Furthermore, based on phylogenetic tree reconstruction, and isolate ST information, resistance did not appear to be associated with a particular clone, except for in the single case of a cluster of serotype 14 ST63 isolates, which displayed similar genetic variation in the *pbp1a* gene. Contrasting the global scenario, which is dominated by a limited number of highly successful pneumococcal lineages (McGee et al., 2001), the Malawian pneumococcal population was also highly mixed in terms of the STs present, and with moderately high rates of carriage, and multiple-carriage (Usuf et al., 2014, Kamng'ona et al., 2015, Everett et al., 2012). This suggests that competition between pneumococcal lineages has perhaps prevented particular lineages form dominating the population here.

The prevalence of meningitis disease in Malawi means that penicillin resistance is frequently viewed in terms of the lower, meningitis resistance breakpoints, resistance classed as an MIC >0.06µg/mL (Weinstein et al., 2009). However, the existence of resistance based on non-meningitis breakpoints (≥2µg/mL)(BSAC, 2013) is comparatively rare (Feikin et al., 2003, Cornick et al., 2011). This, in addition to the diversity in STs present in Malawi suggest that despite the wide availability of penicillin, the usage of this antimicrobial does not appear to have been sufficient to select for high level resistance, and to result in reduced nasopharyngeal carriage.

Interestingly, the introduction of ceftriaxone does appear to lead to a selection for reduced susceptibility to this antimicrobial, despite it's restricted usage, and more recent availability in Malawi (Everett et al., 2011). The high levels of *pbp1a* and *pbp2x* diversity recorded here, indicative of greater rates of recombination occurring among these two genes, compared to *pbp2b* could, in accordance with the above observation, suggest ceftriaxone is providing a selective pressure for modification within these target genes. At present however, ceftriaxone susceptibility information has not been routinely collected at QECH, consequently, this may be an important consideration in the future in order to monitor the emergence of resistance to this antimicrobial in Malawi.

Perhaps of concern will be the effect of continued roll-out of the PCV13 vaccination on the circulating pneumococcal population in Malawi. PCV13 is expected to protect against approximately 60% of all circulating strains, including serotype 1, a prevalent cause of IPD in Africa as a whole (Everett et al., 2012). However, PCV13 may also affect the carriage dynamics of circulating pneumococci in Malawi, potentially favouring MDR lineages, which are often less fit under normal growth conditions owing to their carrying PBP modifications (Rudolf et al., 2011). The occurrence of a common, shared penicillin resistance mechanism among *pbp1a* alleles belonging to ST63 could indicate that this is a lineage that could benefit from a change in the circulating STs. Serotype 14 is included in PCV13, and consequently should be eliminated over the roll out of PCV13. Despite this, Everett and colleagues (2012) have already indicated the potential for currently circulating strains to switch to non-vaccine type serotypes. This highlights the importance for continued genetic analysis in order to monitor the effects of vaccination on the circulating pneumococcal lineages into the future.

8.2 The Future of Resistance in Malawi

The future of beta-lactam resistance in Malawi remains an area of speculation. Given the history of pneumococcal disease, it is likely that ceftriaxone resistance will continue to spread. Higher levels of resistance may also be selected for

owing to the increased risk of treatment breakdowns. However, the roll-out of PCV13 could counteract this process, by removing drug resistant pneumococcal lineages from the population, and in addition by reducing IPD, reduce antibiotic usage, and the selective pressure this provides in Malawi. Conversely, as described above, PCV13 may similarly affect the current competitive balance between pneumococcal lineages, perhaps favouring the spread of particular clones, that may be more associated with drug resistance. Consequently it will be important to continue to monitor the population dynamics of pneumococci during this period of changing selective pressures. This is particularly important given the demonstration that pneumococci, as well as *S. mitis,* and *S. oralis* can circulate on a global scale, so that different pneumococcal lineages could similarly avail of the changes in carriage dynamics in the Malawian pneumococcal population.

Although the global circulation and interaction of these species has been investigated, the inability to identify *S. mitis, S. oralis* and *S. pneumoniae* from nasopharyngeal samples means that the local viridans population remains uncharacterised in terms of antibiotic susceptibility, frequency of co-colonisation, and the nature of the genetic material being exchanged between these species. As such this appears to be an area for future investigation.

Chewapreecha and colleagues (2014a) attempted to identify genetic donors and recipients from a dataset of over 3000 pneumococci isolated from nasopharyngeal samples. These samples were collected from a refugee camp, Maela, on the Thailand-Myanmar border. However, even within this relatively closed population, detecting the donors and recipients of genetic fragments was only achieved by employing a very dense sampling strategy (Chewapreecha et al., 2014a).

As such, preferential targeting of samples in which pneumococci can be isolated alongside *S. mitis* and *S. oralis* is likely to be important in order to maximise the chance of identifying interspecies genetic exchange events. This study has highlighted a number of microbiological methods that could be applied to

achieve this sampling strategy, including the use of Mitis Salivarious Agar. In addition, it is likely that targeting infants for study would be an important consideration.

Although it would be difficult to target *S. mitis* and *S. oralis* for clinical intervention it is still important to understand the role that these species have and continue to play on the development of resistance among pneumococci. For instance, if such species recombine rarely, this could be an important observation in favour of the use of therapies that aim to eradicate pneumococcal carriage- as this would presumably lead to further reductions the opportunities for genetic exchanges to occur between theses species, and so for pneumococci to acquire drug resistance via this route (Bogaert et al., 2004).

8.3 GWAS and the MDR lineage

The GWAS method developed here demonstrates the potential for this method to be applied even to relatively small datasets. This study also acts as a strong comparator to the more complex, and much larger, multiple lineage analysis carried out by Chewapreecha and colleagues (2014b). The single lineage approach offered an important opportunity to assess the mechanisms by which a globally important MDR clone has developed beta-lactam resistance. PMEN1 was found to carry a number of highly conserved SNPs across the dataset, proven *in vitro* to affect beta-lactam susceptibilities. As such, these could represent the core mutations that have allowed PMEN1 to develop resistance.

In addition to PBP modification, several genes were found to carry SNPs statistically associated with a change in beta-lactam susceptibility. The role of MurF and DdlA in beta lactam resistance remains to be investigated *in vitro.* Whilst such associations could result from linkage across the *pbp2b* gene (Smith and Klugman, 1998), no study to date has investigated whether the activity of these genes is affected by the frequent recombination that affects this site. Given the role of these genes in the cell synthesis pathway, it is therefore plausible that

modifications to such genes could affect the downstream interactions between *pbps* and substrates.

The *leuS* gene was also identified as carrying SNPs statistically associated with changes in beta-lactam susceptibilities. This gene plays a role in sequestering of PBPs within the cell (Vinella et al., 1993). Chewapreecha and colleagues (2014b) similarly identified a gene involved in this sequestering process supporting this as a mechanism for affecting beta-lactam susceptibility. The finding that this gene appears to undergo genetic transformation with *S. mitis* and *S. oralis* provided additional support for this hypothesis. However, transformation experiments would be necessary to confirm the proposed relationship between this gene and beta-lactam resistance.

The GWAS method applied here similarly highlights some of the shortcomings of applying this method to the study of bacterial populations. Recombination is an important consideration in this regard. Even when using large-scale bacterial datasets the effect of this process has not been entirely removed (Chewapreecha et al., 2014a). Consequently, it may not be possible to control for the effects of recombination. Whilst *in vitro* testing of genes found to associate with resistance might offer a solution, it is becoming more widely appreciated that the genetic background plays an important role in the level of beta-lactam resistance observed (Chesnel et al., 2005). Consequently, it may be difficult to decipher the precise role of a resistance-associated gene in an isolates' antibiotic susceptibility. Furthermore, *in vitro* testing can be more costly and time consuming than the process of generating a GWAS analysis. As such it is important that candidate genes for testing are chosen with care. In this respect the role of recombination between viridans streptococci and *leuS* could be an important observation, as this potentially provides independent information, that similar to known resistance genes (i.e. the *pbps*), this gene engages in recombination with viridans species.

8.4 Oxidative Stress and Repair

Free radical attack is a problem for all organisms, and as such they must protect their internal genetic material and proteins from such damage. For the pneumococcus, which exists in an oxygenated environment, and produces hydrogen peroxide as part of its metabolism, it is expected to be particularly susceptible to such attack (Yesilkaya et al., 2013). Continuous growth cultures using the sorbarod method have previously documented the occurrence of tandem duplication within the capsule locus, discussed at the time as a mechanism for regulating capsule expression (Waite et al., 2001, Waite et al., 2003). However, tandem duplication is also an artefact of oxidative damage, suggesting that Waite's original observations may in fact result from free radical attack (Pericone et al., 2002).

The sorbarod apparatus allows a period of extended pneumococcal growth to be achieved, and has been used as a model for nasopharyngeal colonisation by the pneumococcus. In contrast, growth under normal laboratory conditions, such as on blood agar plate or in broth, results in the pneumococcus undergoing autolysis after a period of approximately 16 hours of growth (Klein et al., 2006). To exacerbate the affects of peroxide stress, pneumococci were grown over a period of 48 hours in the study presented here, subject to continuous perfusion with hydrogen peroxide infused media.

Despite continuous exposure to millimolar concentrations of hydrogen peroxide over this period, this analysis suggests that the pneumococci tested were surprisingly resilient to attempts to induce oxidative stress in the laboratory. A limited amount of nucleotide variation was observed, notably one large deletion, in addition to several SNPs, which had appeared to result in capsule inactivation. Many of the polymorphisms observed were also limited to repeat regions in the genome- regions that can be repaired intrinsically through slip-strand mispairing (Waite et al., 2001).

Whilst the genome assemblies suggested that variation was limited, in several sites, the sequence reads appeared to be mixed, with some spanning the region,

and some containing an indel. It is unclear whether such regions resulted from mixed sequences present in the original sample, or whether these resulted from sequencing errors. If such reads were truly representative of the original sample this would support the repair hypothesis described *in silico.* Here, healthy strains could act as donors for strains subject to oxidative stress, and as a result, sequencing a lawn of bacteria would result in a mixture of reads derived from fit and damaged pneumococci alike. Current technologies, which suffer from sequencing errors, short read lengths, and require large amounts of starting material therefore inhibit the ability to characterise the diversity present in a sample.

The ability to detect oxidative stress using this method is also subject to unavoidable difficulties. In order to acquire sufficient culture for sequencing, single strains were subject to overnight growth, with up to four plates required. As such, if repair occurs at a high frequency, as predicted from the *in silico* analysis, reversal and repair of genomic damage could have occurred during this growth period. This would therefore lead to an under-representation of the true amount of damage present following oxidative stress. Furthermore, by only being able to sequence viable colonies, additional restrictions were placed on the ability to detect genome-wide oxidative damage. Finally, it is also possible that perhaps due to the short diffusion distances for the potent hydroxyl free radicals, (Redmond and Kochevar, 2006), external delivery of hydrogen peroxide may be a poor method for modelling oxidative stress as it occurs *in vivo.*

The *spxB* gene was found to be highly conserved among PMEN1 isolates. As expected hydrogen peroxide production was similar among these isolates. Interestingly PMEN1 isolates were found to produce higher concentrations of this toxin than the control D39 strain. Given the importance of hydrogen peroxide for the inhibition of co-inhabiting bacteria, and its effect on genetic damage, different abilities for pneumococci to produce this compound, could affect their *in vivo* fitness. Consequently, the expansion of this analysis to include other pneumococcal lineages would be an important future direction for analysis.

Despite the apparent robustness of pneumococci to withstand hydrogen peroxide delivered externally, *in silico,* the deletion detection scheme developed here identified genome-wide damage among a number of clinically sampled pneumococcal lineages. The occurrence of deletions in this analysis was based over years, rather than 48 hours, with such events appearing to occur randomly throughout the bacterial genome, ranging in size from single to several hundred base pairs in length. Consequently, the repair of such damage must be an important mechanism for pneumococci in order to maintain genomic integrity despite free radical attack.

The method developed here allowed the identification of events where daughter strains that had diverged from a common ancestor had lost indels. This process was termed "repair", and allowed recombination events to be identified that could not be detected based on conventional SNP based approaches (Croucher et al., 2015). Building on previous studies that allow the identification of recombination events genome-wide it was further possible to determine whether a recombination event had occurred between a parental strain, or a more distantly related strain, and whether this had lead to the repair of an indel "non-self repair", or whether repair had occurred independently of this "self-repair". This analysis suggested that non-self repair occurs at an almost negligible level when compared to self-repair.

In accordance with this finding, pneumococcal MLST genes display far less genetic diversity when compared to other closely related streptococci (1% compared to ~5% for it's close relatives) (Hanage et al., 2006). If self-repair occurs frequently, this would suggest that pneumococci are able to purge genetic variation through recombination events, helping to maintain these low levels of diversity within their housekeeping genes.

In the absence of such a repair mechanism the continued accumulation of deletions would eventually make the cell unviable. Interestingly, increasing evidence suggests that pneumococci become competent in response to genetic

damage, which would be expected if recombination functions as a mechanism of genetic repair (Bortoni et al., 2009). Self-repair could furthermore help explain the occurrence of fratricide among pneumococci. During this process competent pneumococci release lytic enzymes, killing non-competent pneumococci, allowing them to acquire DNA from these cells (Claverys et al., 2007).

8.5 Indel detection problems

Detecting indels accurately *in situ* however remains problematic. Read length is an important consideration in this respect. Consequently indel detection is likely to improve as the technology develops and longer reads can be accurately sequenced. In addition to longer reads, this study also outlined a number of different approaches to indel detection that were not incorporated into the detection scheme used here. A method for detecting indels in bacterial populations using the split read alignment process has been published previously (Harris et al., 2013a). Other methods reliant on changes in the internal segment size, or differences in mapping coverage however have not been as widely applied (Marschall et al., 2013). Furthermore, coverage based approaches may suffer from particular regions of the genome inherently having low mapping coverage- such as due to repetition elsewhere in the genome. Consequently additional quality control methods may need to be developed in order to include these methods of indel detection into the current protocol.

Whilst many methods have been developed for indel detection, these are all designed to overcome a fundamental flaw in the sequencing process- short reads. Consequently, there is a growing need for longer reads, of high accuracy. PacBio is the current leader in this field, with read length in the region of kilobases, however, this technology remains too expensive to be used as the default method for genome sequencing (Quail et al., 2012). The amount of sample required for sequencing is a further limitation as discussed previously. The *in vitro* repair, or formation of deletion as a result of preparing sufficient quantities of sample for sequencing will affect the ability to measure the occurrence of such events *in vivo*. One laboratory practice that could reduce the

occurrence of these processes would be to grow pneumococci under anaerobic conditions, so that their metabolism does not result in the production of hydrogen peroxide, likely largely responsible for such genetic damage (Carvalho et al., 2013).

The sampling strategy and reference chosen during mapping will similarly affect the ability to detect indels. It is important that the reference sequence chosen belongs to a close lineage if a member of the same ST is not available for analysis. Otherwise, sequence divergence between the reference and study population can lead to assembly errors, which can mistakenly result in the detection of indels. Furthermore, insertions and deletions may be present between the reference sequence and the study population. In this case indels will be over-represented in the assembled reads.

Repetitive sequence represents a further complication for accurate indel detection. Changes in the length of repeat regions were detected in isolates subject to oxidative stress *in vitro* (chapter 6), and during the analysis of clinical samples (chapter 7). Gene redundancy, such as in the case of the 16S genes can also lead to larger numbers of reads mapping to the wrong locations within the genome. As such false "indels" can be detected. The method described in section 7 relied on removing repetitive sequences, and an upper limit on indel detection was also chosen to remove spurious indels. However, due to the low GC content of the pneumococcal genome, it is likely that short indels comprising A and T nucleotides are likely to be removed more frequently. As such, the results presented here are likely to be an under-representation of the true number of indels present. In addition, longer read length would improve this inaccuracy.

A final consideration is the effect of sampling strategy. Whereas PMEN isolates were studied globally, samples chosen from Thailand were sampled over a much shorter time frame. Such dense sampling strategies appear to improve the ability to detect repair events. This is likely because branches of the phylogeny are better represented when dense sampling is undertaken, compared to when samples are collected from disparate populations globally.

8.6 Future repair analyses

The potential importance of self-repair suggests that this process may occur more broadly. As such it will be important to identify whether this process can be found within other bacterial species. Self-repair could be hypothesised to be particularly important for intracellular pathogens, exposed to substantial free radical attack from the host's immune system during colonisation. In this respect pathogens such as *N. gonorrhoeae, Chlamydia trachomatis* and *Helicobacter pylori* for which large genomic datasets for clinical isolates already exist, would be particularly appealing for future study (Grad et al., 2014, Harris et al., 2012, Dong et al., 2014). Interestingly, although traditionally considered clonal, there is growing evidence to suggest recombination does occur in some of these species. Consequently, self-repair, as a result of inhabiting an oxygen-stressed environment may be an important mechanism allowing such species to maintain genomic fidelity.

In addition, particular conditions can place organisms under high levels of oxidative stress. Sickle cell disease is a particularly important syndrome in this respect as sickle cell haemoglobins produce significantly increased amounts of free radicals compared to normal red blood cells (Gizi et al., 2011). Consequently, pathogens that invade the bloodstream of sickle cell sufferers will become subject to much higher rates of oxidative stress than are normally present. This could be expected to increase genomic damage, and as such the occurrence of recombination in such isolates.

The role that phage and other genomic variation has in inactivating *com* genes is also of interest. Given the implication for the importance of recombination in genome repair, inactivation of these genes would be expected to be particularly detrimental to the host. The precise benefit of disrupting competence genes by bacteriophages has not been investigated. However, the occurrence of this phenomenon is becoming more widely documented (Croucher et al., 2011, Croucher et al., 2014b, Harris et al., 2015). Consequently, some benefit of restricting host recombination may be conferred, perhaps to facilitate further phage invasion.

In many bacterial species deletions are marked by specific motif regions, represented by either a specific base pattern, or a particular base prevalence. In *E. coli* Chi sites occur at a high frequency throughout the genome and play an important role in the repair of double-strand breaks through their interaction with RecBCD (Smith, 2012). As such, in the same way that Chi sequences can be an indication of genetic repair in *E. coli* the occurrence of particular motifs might be expected to occur near to the deletions in pneumococci (El Karoui et al., 1999, Smith, 2012). Currently no such sites are known to exist in pneumococci, and no such sites were detected in this analysis. As such there appears to be deficit in knowledge as to the mode by which deletions are formed and how non-homologous end joining is regulated in pneumococcus.

Finally, whilst deletions have been identified widely *in silico* it will be important to verify these occurrences *in vitro.* This will be particularly important in cases where large indels are detected, which exceed read length (Marschall et al., 2013).

8.7 Final summary

In final summary, transformation appears to play a role not only in the development of beta-lactam resistance but in the repair of genetic damage, inflicted, likely as a result of oxidative stress *in vivo*. Deletion repair more broadly suggest bacteria could engage in recombination far more frequently than previously thought. This could have far reaching evolutionary consequences, such as in the emergence of drug resistance, and the survival of pathogens in their hosts. Furthermore, this provides important evidence for the role of recombination as a mechanism of genome repair. The above analyses have also demonstrated that beta-lactam resistance remains in development in Malawi, and should be monitored in terms of the role that *S. mitis* and *S. oralis* have in this process, and the role that globally circulating streptococcal species similarly play in the dissemination of beta-lactam resistance.

9 References

AANENSEN, D. M., MAVROIDI, A., BENTLEY, S. D., REEVES, P. R. & SPRATT, B. G. 2007. Predicted functions and linkage specificities of the products of the Streptococcus pneumoniae capsular biosynthetic loci. *J Bacteriol,* 189**,** 7856-76.

ABEYTA, M. 1999. *thesis.* University of Alabama.

ABRAHAM, E. P., CHAIN, E., FLETCHER, C. M., GARDNER, A. D., HEATLEY, N. G., JENNINGS, M. A. & FLOREY, H. W. 1941. FURTHER OBSERVATIONS ON PENICILLIN. *The Lancet,* 238**,** 177-189.

ABYZOV, A., URBAN, A. E., SNYDER, M. & GERSTEIN, M. 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res,* 21**,** 974-84.

ADAMS, W. G., DEAVER, K. A., COCHI, S. L., PLIKAYTIS, B. D., ZELL, E. R., BROOME, C. V. & WENGER, J. D. 1993. Decline of childhood Haemophilus influenzae type b (Hib) disease in the Hib vaccine era. *Jama,* 269**,** 221-6.

ADEGBOLA, R. A., HILL, P. C., SECKA, O., IKUMAPAYI, U. N., LAHAI, G., GREENWOOD, B. M. & CORRAH, T. 2006. Serotype and antimicrobial susceptibility patterns of isolates of Streptococcus pneumoniae causing invasive disease in The Gambia 1996-2003. *Trop Med Int Health,* 11**,** 1128-35.

AHL, J., LITTORIN, N., FORSGREN, A., ODENHOLT, I., RESMAN, F. & RIESBECK, K. 2013. High incidence of septic shock caused by Streptococcus pneumoniae serotype 3--a retrospective epidemiological study. *BMC Infect Dis,* 13**,** 492.

AHMAN, H., KAYHTY, H., TAMMINEN, P., VUORELA, A., MALINOSKI, F. & ESKOLA, J. 1996. Pentavalent pneumococcal oligosaccharide conjugate vaccine PncCRM is well-tolerated and able to induce an antibody response in infants. *Pediatr Infect Dis J,* 15**,** 134-9.

ALBRICH, W. C., MADHI, S. A., ADRIAN, P. V., VAN NIEKERK, N., TELLES, J. N., EBRAHIM, N., MESSAOUDI, M., PARANHOS-BACCALA, G., GIERSDORF, S., VERNET, G., MUELLER, B. & KLUGMAN, K. P. 2014. Pneumococcal colonisation density: a new marker for disease severity in HIV-infected adults with pneumonia. *BMJ Open,* 4**,** e005953.

ALBRICH, W. C., MONNET, D. L. & HARBARTH, S. 2004. Antibiotic selection pressure and resistance in Streptococcus pneumoniae and Streptococcus pyogenes. *Emerg Infect Dis,* 10**,** 514-7.

ALLEGRUCCI, M. & SAUER, K. 2007. Characterization of colony morphology variants isolated from Streptococcus pneumoniae biofilms. *J Bacteriol,* 189**,** 2030-8.

ALONSO, J. C., CARDENAS, P. P., SANCHEZ, H., HEJNA, J., SUZUKI, Y. & TAKEYASU, K. 2013. Early steps of double-strand break repair in Bacillus subtilis. *DNA Repair (Amst),* 12**,** 162-76.

ALOUF, J. E. 2000. Cholesterol-binding cytolytic protein toxins. *International Journal of Medical Microbiology,* 290**,** 351-356.

ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res,* 25**,** 3389-402.

AMOROSO, A., DEMARES, D., MOLLERACH, M., GUTKIND, G. & COYETTE, J. 2001. All Detectable High-Molecular-Mass Penicillin-Binding Proteins Are Modified in a High-Level β-Lactam-Resistant Clinical Isolate of Streptococcus mitis. *Antimicrob Agents Chemother,* 45**,** 2075-81.

ANDISI, V. F., HINOJOSA, C. A., DE JONG, A., KUIPERS, O. P., ORIHUELA, C. J. & BIJLSMA, J. J. 2012. Pneumococcal gene complex involved in resistance to extracellular oxidative stress. *Infect Immun,* 80**,** 1037-49.

ANIANSSON, G., ALM, B., ANDERSSON, B., LARSSON, P., NYLEN, O., PETERSON, H., RIGNER, P., SVANBORG, M. & SVANBORG, C. 1992. Nasopharyngeal colonization during the first year of life. *J Infect Dis,* 165 Suppl 1**,** S38-42.

ANTTILA, M., ESKOLA, J., ÅHMAN, H. & KÄYHTY, H. 1999. Differences in the avidity of antibodies evoked by four different pneumococcal conjugate vaccines in early childhood1. *Vaccine,* 17**,** 1970-1977.

APPELBAUM, P. C. 1987. World-wide development of antibiotic resistance in pneumococci. *European Journal of Clinical Microbiology,* 6**,** 367-377.

APPELBAUM, P. C. 1992. Antimicrobial Resistance in Streptococcus pneumoniae: An Overview. *Clinical Infectious Diseases,* 15**,** 77-83.

ARBIQUE, J. C., POYART, C., TRIEU-CUOT, P., QUESNE, G., CARVALHO MDA, G., STEIGERWALT, A. G., MOREY, R. E., JACKSON, D., DAVIDSON, R. J. & FACKLAM, R. R. 2004. Accuracy of phenotypic and genotypic testing for identification of Streptococcus pneumoniae and description of Streptococcus pseudopneumoniae sp. nov. *J Clin Microbiol,* 42**,** 4686-96.

ARRECUBIETA, P. C., GLADKOVA, C. & HRYNIEWICZ, W. O. A. 1996. Type 3-specific synthase of Streptococcus pneumoniae (Cap3B) directs type 3 polysaccharide biosynthesis in Escherichia coli and in pneumococcal strains of different serotypes. *J Exp Med,* 184**,** 449-55.

ATTAIECH, L., OLIVIER, A., MORTIER-BARRIÈRE, I., SOULET, A.-L., GRANADEL, C., MARTIN, B., POLARD, P. & CLAVERYS, J.-P. 2011. Role of the Single-Stranded DNA–Binding Protein SsbB in Pneumococcal Transformation: Maintenance of a Reservoir for Genetic Plasticity. *PLoS Genet,* 7**,** e1002156.

AU, N., KUESTER-SCHOECK, E., MANDAVA, V., BOTHWELL, L. E., CANNY, S. P., CHACHU, K., COLAVITO, S. A., FULLER, S. N., GROBAN, E. S., HENSLEY, L. A., O'BRIEN, T. C., SHAH, A., TIERNEY, J. T., TOMM, L. L., O'GARA, T. M., GORANOV, A. I., GROSSMAN, A. D. & LOVETT, C. M. 2005. Genetic composition of the Bacillus subtilis SOS system. *J Bacteriol,* 187**,** 7655-66.

AUSTRIAN, R. 1978. *Of Gold and Pneumococci: A History of Pneumococcal Vaccines in South Africa.*

AUSTRIAN, R. 1999. The pneumococcus at the millennium: not down, not out. *J Infect Dis,* 179 Suppl 2**,** S338-41.

AUSTRIAN, R., DOUGLAS, R. M., SCHIFFMAN, G., COETZEE, A. M., KOORNHOF, H. J., HAYDEN-SMITH, S. & REID, R. D. 1976. Prevention of pneumococcal pneumonia by vaccination. *Trans Assoc Am Physicians,* 89**,** 184-94.

AUSTRIAN, R. & GOLD, J. 1964. Pneumococcal Bacteremia with Especial Reference to Bacteremic Pneumococcal Pneumonia. *Annals of Internal Medicine,* 60**,** 759-776.

AUSTRIAN, R. & ROSENBLUM, R. 1953. The relative efficacy of erythromycin (ilotycin) and of penicillin in the treatment of pneumococcal lobar pneumonia. *Am J Med Sci,* 226**,** 487-90.
AUZAT, I., CHAPUY-REGAUD, S., LE BRAS, G., DOS SANTOS, D., OGUNNIYI, A. D., LE THOMAS, I., GAREL, J. R., PATON, J. C. & TROMBE, M. C. 1999. The NADH oxidase of Streptococcus pneumoniae: its involvement in competence and virulence. *Mol Microbiol,* 34**,** 1018-28.
AVERY, O. T., CHICKERING, H. T. & DOCHEZ, A. R. 1917. Acute lobar pneumonia, prevention and serum treatment. *Monographs of the Rockefeller Institute,* 7**,** 1-110.
AVERY, O. T. & GOEBEL, W. F. 1929. Chemo-immunological studies on conjugated carbohydrate-proteins: II. Immunological specificity of synthetic sugar-protein antigens. *J Exp Med,* 50**,** 533-50.
AVERY, O. T., MACLEOD, C. M. & MCCARTY, M. 1944. Studies on the chemical nature of the substances inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *The Journal of Experimental Medicine,* 79**,** 137-158.
AYORA, S., CARRASCO, B., CARDENAS, P. P., CESAR, C. E., CANAS, C., YADAV, T., MARCHISONE, C. & ALONSO, J. C. 2011. Double-strand break repair in bacteria: a view from Bacillus subtilis. *FEMS Microbiol Rev,* 35**,** 1055-81.
BAHAROGLU, Z. & MAZEL, D. 2014. SOS, the formidable strategy of bacteria against aggressions. *FEMS Microbiol Rev,* 38**,** 1126-45.
BAKHSHAEE, M., NADERI, H. R., GHAZVINI, K., SOTOUDEH, K., AMALI, A. & ASHTIANI, S. J. 2012. Passive smoking and nasopharyngeal colonization by Streptococcus pneumoniae, Haemophilus influenzae, and Moraxella catarrhalis in daycare children. *Eur Arch Otorhinolaryngol,* 269**,** 1127-32.
BALABAN, M., BÄTTIG, P., MUSCHIOL, S., TIRIER, S. M., WARTHA, F., NORMARK, S. & HENRIQUES-NORMARK, B. 2014. Secretion of a pneumococcal type II secretion system pilus correlates with DNA uptake during transformation. *Proceedings of the National Academy of Sciences,* 111**,** E758-E765.
BAQUERO, F., BAQUERO-ARTIGAO, G., CANTON, R. & GARCIA-REY, C. 2002. Antibiotic consumption and resistance selection in Streptococcus pneumoniae. *J Antimicrob Chemother,* 50 Suppl S2**,** 27-37.
BARBOUR, M. L., MAYON-WHITE, R. T., COLES, C., CROOK, D. W. & MOXON, E. R. 1995. The impact of conjugate vaccine on carriage of Haemophilus influenzae type b. *J Infect Dis,* 171**,** 93-8.
BARCUS, V. A., GHANEKAR, K., YEO, M., COFFEY, T. J. & DOWSON, C. G. 1995. Genetics of high level penicillin resistance in clinical isolates of Streptococcus pneumoniae. *FEMS Microbiol Lett,* 126**,** 299-303.
BARRETT-CONNOR, E. 1971. Bacterial infection and sickle cell anemia. An analysis of 250 infections in 166 patients and a review of the literature. *Medicine (Baltimore),* 50**,** 97-112.
BARYSHNIKOVA, A., COSTANZO, M., DIXON, S., VIZEACOUMAR, F. J., MYERS, C. L., ANDREWS, B. & BOONE, C. 2010. Synthetic genetic array (SGA) analysis in Saccharomyces cerevisiae and Schizosaccharomyces pombe. *Methods Enzymol,* 470**,** 145-79.
BEALL, B., MCELLISTREM, M. C., GERTZ, R. E., JR., WEDEL, S., BOXRUD, D. J., GONZALEZ, A. L., MEDINA, M. J., PAI, R., THOMPSON, T. A., HARRISON, L.

H., MCGEE, L. & WHITNEY, C. G. 2006. Pre- and postvaccination clonal compositions of invasive pneumococcal serotypes for isolates collected in the United States in 1999, 2001, and 2002. *J Clin Microbiol,* 44**,** 999-1017.

BEESON, P. B. & GOEBEL, W. F. 1939. The immunological relationship of the capsular polysaccharide of type XIV pneumococcus to the blood group A specific substance. *J Exp Med,* 70**,** 239-47.

BENTLEY, D. R., BALASUBRAMANIAN, S., SWERDLOW, H. P., SMITH, G. P., MILTON, J., BROWN, C. G., HALL, K. P., EVERS, D. J., BARNES, C. L., BIGNELL, H. R., BOUTELL, J. M., BRYANT, J., CARTER, R. J., KEIRA CHEETHAM, R., COX, A. J., ELLIS, D. J., FLATBUSH, M. R., GORMLEY, N. A., HUMPHRAY, S. J., IRVING, L. J., KARBELASHVILI, M. S., KIRK, S. M., LI, H., LIU, X., MAISINGER, K. S., MURRAY, L. J., OBRADOVIC, B., OST, T., PARKINSON, M. L., PRATT, M. R., RASOLONJATOVO, I. M., REED, M. T., RIGATTI, R., RODIGHIERO, C., ROSS, M. T., SABOT, A., SANKAR, S. V., SCALLY, A., SCHROTH, G. P., SMITH, M. E., SMITH, V. P., SPIRIDOU, A., TORRANCE, P. E., TZONEV, S. S., VERMAAS, E. H., WALTER, K., WU, X., ZHANG, L., ALAM, M. D., ANASTASI, C., ANIEBO, I. C., BAILEY, D. M., BANCARZ, I. R., BANERJEE, S., BARBOUR, S. G., BAYBAYAN, P. A., BENOIT, V. A., BENSON, K. F., BEVIS, C., BLACK, P. J., BOODHUN, A., BRENNAN, J. S., BRIDGHAM, J. A., BROWN, R. C., BROWN, A. A., BUERMANN, D. H., BUNDU, A. A., BURROWS, J. C., CARTER, N. P., CASTILLO, N., CHIARA, E. C. M., CHANG, S., NEIL COOLEY, R., CRAKE, N. R., DADA, O. O., DIAKOUMAKOS, K. D., DOMINGUEZ-FERNANDEZ, B., EARNSHAW, D. J., EGBUJOR, U. C., ELMORE, D. W., ETCHIN, S. S., EWAN, M. R., FEDURCO, M., FRASER, L. J., FUENTES FAJARDO, K. V., SCOTT FUREY, W., GEORGE, D., GIETZEN, K. J., GODDARD, C. P., GOLDA, G. S., GRANIERI, P. A., GREEN, D. E., GUSTAFSON, D. L., HANSEN, N. F., HARNISH, K., HAUDENSCHILD, C. D., HEYER, N. I., HIMS, M. M., HO, J. T., HORGAN, A. M., et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature,* 456**,** 53-9.

BENTLEY, S. D., AANENSEN, D. M., MAVROIDI, A., SAUNDERS, D., RABBINOWITSCH, E., COLLINS, M., DONOHOE, K., HARRIS, D., MURPHY, L., QUAIL, M. A., SAMUEL, G., SKOVSTED, I. C., KALTOFT, M. S., BARRELL, B., REEVES, P. R., PARKHILL, J. & SPRATT, B. G. 2006. Genetic Analysis of the Capsular Biosynthetic Locus from All 90 Pneumococcal Serotypes. *PLoS Genet,* 2**,** e31.

BERG, K. H., STAMSAS, G. A., STRAUME, D. & HAVARSTEIN, L. S. 2013. Effects of low PBP2b levels on cell morphology and peptidoglycan composition in Streptococcus pneumoniae R6. *J Bacteriol,* 195**,** 4342-54.

BERGE, M., MORTIER-BARRIERE, I., MARTIN, B. & CLAVERYS, J. P. 2003. Transformation of Streptococcus pneumoniae relies on DprA- and RecA-dependent protection of incoming DNA single strands. *Mol Microbiol,* 50**,** 527-36.

BERGÉ, M. J., KAMGOUÉ, A., MARTIN, B., POLARD, P., CAMPO, N. & CLAVERYS, J.-P. 2013. Midcell Recruitment of the DNA Uptake and Virulence Nuclease, EndA, for Pneumococcal Transformation. *PLoS Pathog,* 9**,** e1003596.

BERRY, A. M., LOCK, R. A. & PATON, J. C. 1996. Cloning and characterization of nanB, a second Streptococcus pneumoniae neuraminidase gene, and

purification of the NanB enzyme from recombinant Escherichia coli. *J Bacteriol,* 178**,** 4854-60.
BEZANÇON, F. & GRIFFON, V. 1897. Pouvoir agglutinatif du serum dans les infections experimentales et humaines a pneumocoques. Part 1. *Comptes rendus des séances de la Société de biologie et de ses filiales,* 49.
BISAILLON, J. G., DUBOIS, G., BEAUDET, R., SYLVESTRE, M., CHARBONNEAU, R. & GAGNON, M. 1985. Quantitative determination of catalase activity produced by Neisseria gonorrhoeae, Staphylococcus epidermidis, Neisseria meningitidis and other bacterial strains using the Catalasemeter. *Exp Biol,* 43**,** 225-30.
BLUESTONE, C. D., STEPHENSON, J. S. & MARTIN, L. M. 1992. Ten-year review of otitis media pathogens. *Pediatr Infect Dis J,* 11**,** S7-11.
BOBBA, S. & GUTHEIL, W. G. 2011. Multivariate geometrical analysis of catalytic residues in the penicillin-binding proteins. *Int J Biochem Cell Biol,* 43**,** 1490-9.
BOETZER, M. & PIROVANO, W. 2012. Toward almost closed genomes with GapFiller. *Genome Biol,* 13**,** R56.
BOETZER, M. & PIROVANO, W. 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics,* 15**,** 211.
BOGAERT, D., DE GROOT, R. & HERMANS, P. W. 2004. Streptococcus pneumoniae colonisation: the key to pneumococcal disease. *Lancet Infect Dis,* 4**,** 144-54.
BOGAERT, D., KEIJSER, B., HUSE, S., ROSSEN, J., VEENHOVEN, R., VAN GILS, E., BRUIN, J., MONTIJN, R., BONTEN, M. & SANDERS, E. 2011. Variability and Diversity of Nasopharyngeal Microbiota in Children: A Metagenomic Analysis. *PLoS ONE,* 6**,** e17035.
BOREK, A. P., DRESSEL, D. C., HUSSONG, J. & PETERSON, L. R. 1997. Evolving clinical problems with Streptococcus pneumoniae: increasing resistance to antimicrobial agents, and failure of traditional optochin identification in Chicago, Illinois, between 1993 and 1996. *Diagn Microbiol Infect Dis,* 29**,** 209-14.
BORST, L. B., PATTERSON, S. K., LANKA, S., SUYEMOTO, M. M. & MADDOX, C. W. 2013. Zebrafish (Danio rerio) as a screen for attenuation of Lancefield group C streptococci and a model for streptococcal pathogenesis. *Vet Pathol,* 50**,** 457-67.
BORTONI, M. E., TERRA, V. S., HINDS, J., ANDREW, P. W. & YESILKAYA, H. 2009. The pneumococcal response to oxidative stress includes a role for Rgg. *Microbiology,* 155**,** 4123-34.
BOWATER, R. & DOHERTY, A. J. 2006. Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS Genet,* 2**,** e8.
BREUKINK, E. & DE KRUIJFF, B. 2006. Lipid II as a target for antibiotics. *Nat Rev Drug Discov,* 5**,** 321-32.
BROOK, I. & GOBER, A. E. 2005. Long-term effects on the nasopharyngeal flora of children following antimicrobial therapy of acute otitis media with cefdinir or amoxycillin-clavulanate. *Journal of Medical Microbiology,* 54**,** 553-556.
BRUEGGEMANN, A. B., GRIFFITHS, D. T., MEATS, E., PETO, T., CROOK, D. W. & SPRATT, B. G. 2003. Clonal relationships between invasive and carriage

Streptococcus pneumoniae and serotype- and clone-specific differences in invasive disease potential. *J Infect Dis,* 187**,** 1424-32.
BRUEGGEMANN, A. B. & SPRATT, B. G. 2003. Geographic Distribution and Clonal Diversity of Streptococcus pneumoniae Serotype 1 Isolates. *Journal of Clinical Microbiology,* 41**,** 4966-4970.
BSAC. 2013. *BSAC Methods for Antimicrobial Susceptiblity Testing* [Online]. Available: http://bsac.org.uk/wp-content/uploads/2012/02/Version-12-Apr-2013_final.pdf [Accessed 8 July 2015].
BSAC. 2014. *The British Society for Antimicrobial Chemotherapy* [Online]. Available: http://bsac.org.uk/wp-content/uploads/2014/06/BSAC-disc-susceptibility-testing-method-June-2014.pdf [Accessed 7 July 2015].
BSAC. 2015. *The British Society for Antimicrobial Chemotherapy* [Online]. Available: http://bsac.org.uk/ [Accessed 7 July 2015].
BUDHANI, R. K. & STRUTHERS, J. K. 1998. Interaction of Streptococcus pneumoniae and Moraxella catarrhalis: Investigation of the Indirect Pathogenic Role of β-Lactamase-Producing Moraxellae by Use of a Continuous-Culture Biofilm System. *Antimicrob Agents Chemother,* 42**,** 2521-6.
BUXTON, R. 2005. *Blood Agar Plates and Hemolysi Protocols* [Online]. Available: http://www.microbelibrary.org/component/resource/laboratory-test/2885-blood-agarplates-and-hemolysis-protocols [Accessed Sept 5th 2015].
CAMPOS, M., CISNEROS, D. A., NIVASKUMAR, M. & FRANCETIC, O. 2013. The type II secretion system - a dynamic fiber assembly nanomachine. *Res Microbiol,* 164**,** 545-55.
CARAPITO, R., CHESNEL, L., VERNET, T. & ZAPUN, A. 2006. Pneumococcal beta-lactam resistance due to a conformational change in penicillin-binding protein 2x. *J Biol Chem,* 281**,** 1771-7.
CARDOZO, D. M., NASCIMENTO-CARVALHO, C. M., ANDRADE, A. L., SILVANY-NETO, A. M., DALTRO, C. H., BRANDAO, M. A., BRANDAO, A. P. & BRANDILEONE, M. C. 2008. Prevalence and risk factors for nasopharyngeal carriage of Streptococcus pneumoniae among adolescents. *J Med Microbiol,* 57**,** 185-9.
CARROL, E. D., GUIVER, M., NKHOMA, S., MANKHAMBO, L. A., MARSH, J., BALMER, P., BANDA, D. L., JEFFERS, G., WHITE, S. A., MOLYNEUX, E. M., MOLYNEUX, M. E., SMYTH, R. L. & HART, C. A. 2007. High pneumococcal DNA loads are associated with mortality in Malawian children with invasive pneumococcal disease. *Pediatr Infect Dis J,* 26**,** 416-22.
CARTER, R., WOLF, J., VAN OPIJNEN, T., MULLER, M., OBERT, C., BURNHAM, C., MANN, B., LI, Y., HAYDEN, R. T., PESTINA, T., PERSONS, D., CAMILLI, A., FLYNN, P. M., TUOMANEN, E. I. & ROSCH, J. W. 2014. Genomic analyses of pneumococci from children with sickle cell disease expose host-specific bacterial adaptations and deficits in current interventions. *Cell Host Microbe,* 15**,** 587-99.
CARVALHO, M. G., STEIGERWALT, A. G., THOMPSON, T., JACKSON, D. & FACKLAM, R. R. 2003. Confirmation of nontypeable Streptococcus pneumoniae-like organisms isolated from outbreaks of epidemic conjunctivitis as Streptococcus pneumoniae. *J Clin Microbiol,* 41**,** 4415-7.

CARVALHO, S. M., FARSHCHI ANDISI, V., GRADSTEDT, H., NEEF, J., KUIPERS, O. P., NEVES, A. R. & BIJLSMA, J. J. 2013. Pyruvate oxidase influences the sugar utilization pattern and capsule production in Streptococcus pneumoniae. *PLoS One,* 8**,** e68277.

CARVER, T. J., RUTHERFORD, K. M., BERRIMAN, M., RAJANDREAM, M. A., BARRELL, B. G. & PARKHILL, J. 2005. ACT: the Artemis Comparison Tool. *Bioinformatics,* 21**,** 3422-3.

CDC. 1996a. *Defining the Public Health Impact of Drug-Resistant Streptococcus pneumoniae: Report of a Working Group* [Online]. Available: http://www.cdc.gov/mmwr/preview/mmwrhtml/00040449.htm [Accessed Sept 4th 2015].

CDC 1996b. Progress toward elimination of Haemophilus influenzae type b disease among infants and children--United States, 1987-1995. *MMWR Morb Mortal Wkly Rep,* 45**,** 901-6.

CDC. 2011. *Laboratory Methods for the Diagnosis of Meningitis* [Online]. Available: http://www.cdc.gov/meningitis/lab-manual/ [Accessed 7 July 2014].

CDC. 2012. *Chapter 8: Identification and Characterisation of Streptococcus pneumoniae* [Online]. Available: http://www.cdc.gov/meningitis/lab-manual/chpt08-id-characterization-streppneumo.html [Accessed 6 July 2015].

CHAIN, E., FLOREY, H. W., ADELAIDE, M. B., GARDNER, A. D., HEATLEY, N. G., JENNINGS, M. A., ORR-EWING, J. & SANDERS, A. G. 1940. Penicillin as a chemotherapeutic agent. *Lancet,* ii**,** 226-228.

CHAKRAVORTY, S., HELB, D., BURDAY, M., CONNELL, N. & ALLAND, D. 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods,* 69**,** 330-9.

CHAPMAN, G. H. 1946. The isolation and testing of fecal streptococci. *Am J Dig Dis,* 13**,** 105-7.

CHEN, D. K., MCGEER, A., DE AZAVEDO, J. C. & LOW, D. E. 1999. Decreased Susceptibility of Streptococcus pneumoniae to Fluoroquinolones in Canada. *New England Journal of Medicine,* 341**,** 233-239.

CHEN, F. M., BREIMAN, R. F., FARLEY, M., PLIKAYTIS, B., DEAVER, K. & CETRON, M. S. 1998. Geocoding and linking data from population-based surveillance and the US Census to evaluate the impact of median household income on the epidemiology of invasive Streptococcus pneumoniae infections. *Am J Epidemiol,* 148**,** 1212-8.

CHEN, K., WALLIS, J. W., MCLELLAN, M. D., LARSON, D. E., KALICKI, J. M., POHL, C. S., MCGRATH, S. D., WENDL, M. C., ZHANG, Q., LOCKE, D. P., SHI, X., FULTON, R. S., LEY, T. J., WILSON, R. K., DING, L. & MARDIS, E. R. 2009. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods,* 6**,** 677-81.

CHEN, Z., YANG, H. & PAVLETICH, N. P. 2008. Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature,* 453**,** 489-4.

CHESNEL, L., CARAPITO, R., CROIZE, J., DIDEBERG, O., VERNET, T. & ZAPUN, A. 2005. Identical penicillin-binding domains in penicillin-binding proteins of Streptococcus pneumoniae clinical isolates with different levels of beta-lactam resistance. *Antimicrob Agents Chemother,* 49**,** 2895-902.

CHEWAPREECHA, C., HARRIS, S. R., CROUCHER, N. J., TURNER, C., MARTTINEN, P., CHENG, L., PESSIA, A., AANENSEN, D. M., MATHER, A. E., PAGE, A. J., SALTER, S. J., HARRIS, D., NOSTEN, F., GOLDBLATT, D., CORANDER, J., PARKHILL, J., TURNER, P. & BENTLEY, S. D. 2014a. Dense genomic sampling identifies highways of pneumococcal recombination. *Nat Genet,* 46**,** 305-9.

CHEWAPREECHA, C., MARTTINEN, P., CROUCHER, N. J., SALTER, S. J., HARRIS, S. R., MATHER, A. E., HANAGE, W. P., GOLDBLATT, D., NOSTEN, F. H., TURNER, C., TURNER, P., BENTLEY, S. D. & PARKHILL, J. 2014b. Comprehensive Identification of Single Nucleotide Polymorphisms Associated with Beta-lactam Resistance within Pneumococcal Mosaic Genes. *PLoS Genet,* 10**,** e1004547.

CHI, F., NOLTE, O., BERGMANN, C., IP, M. & HAKENBECK, R. 2007. Crossing the barrier: evolution and spread of a major class of mosaic pbp2x in Streptococcus pneumoniae, S. mitis and S. oralis. *Int J Med Microbiol,* 297**,** 503-12.

CHIEN, Y.-W., LEVIN, B. R. & KLUGMAN, K. P. 2012. The Anticipated Severity of a "1918-Like" Influenza Pandemic in Contemporary Populations: The Contribution of Antibacterial Interventions. *PLoS ONE,* 7**,** e29219.

CHIEN, Y. W., KLUGMAN, K. P. & MORENS, D. M. 2010. Efficacy of whole-cell killed bacterial vaccines in preventing pneumonia and death during the 1918 influenza pandemic. *J Infect Dis,* 202**,** 1639-48.

CHRISTMAN, M. F., STORZ, G. & AMES, B. N. 1989. OxyR, a positive regulator of hydrogen peroxide-inducible genes in Escherichia coli and Salmonella typhimurium, is homologous to a family of bacterial regulatory proteins. *Proc Natl Acad Sci U S A,* 86**,** 3484-8.

CIMA-CABAL, M. D., VAZQUEZ, F., DE LOS TOYOS, J. R. & MENDEZ, F. J. 1999. Rapid and reliable identification of Streptococcus pneumoniae isolates by pneumolysin-mediated agglutination. *J Clin Microbiol,* 37**,** 1964-6.

CIRZ, R. T., JONES, M. B., GINGLES, N. A., MINOGUE, T. D., JARRAHI, B., PETERSON, S. N. & ROMESBERG, F. E. 2007. Complete and SOS-mediated response of Staphylococcus aureus to the antibiotic ciprofloxacin. *J Bacteriol,* 189**,** 531-9.

CISNEROS, D. A., PEHAU-ARNAUDET, G. & FRANCETIC, O. 2012. Heterologous assembly of type IV pili by a type II secretion system reveals the role of minor pilins in assembly initiation. *Mol Microbiol,* 86**,** 805-18.

CLARIDGE, J. E., 3RD, ATTORRI, S., MUSHER, D. M., HEBERT, J. & DUNBAR, S. 2001. Streptococcus intermedius, Streptococcus constellatus, and Streptococcus anginosus ("Streptococcus milleri group") are of different clinical importance and are not equally associated with abscess. *Clin Infect Dis,* 32**,** 1511-5.

CLAVERYS, J. P. & HAVARSTEIN, L. S. 2007. Cannibalism and fratricide: mechanisms and raisons d'etre. *Nat Rev Microbiol,* 5**,** 219-29.

CLAVERYS, J. P., MARTIN, B. & HAVARSTEIN, L. S. 2007. Competence-induced fratricide in streptococci. *Mol Microbiol,* 64**,** 1423-33.

CLAVERYS, J. P., MARTIN, B. & POLARD, P. 2009. The genetic transformation machinery: composition, localization, and mechanism. *FEMS Microbiol Rev,* 33**,** 643-56.

CLAVERYS, J. P., PRUDHOMME, M. & MARTIN, B. 2006. Induction of competence regulons as a general response to stress in gram-positive bacteria. *Annu Rev Microbiol,* 60**,** 451-75.
CLAVERYS, J. P., PRUDHOMME, M., MORTIER-BARRIERE, I. & MARTIN, B. 2000. Adaptation to the environment: Streptococcus pneumoniae, a paradigm for recombination-mediated genetic plasticity? *Mol Microbiol,* 35**,** 251-9.
CLSI. 2015. *Clinical Laboratory Standards Institute* [Online]. Available: http://clsi.org/ [Accessed 7 July 2015].
COFFEY, T. J., BERRON, S., DANIELS, M., GARCIA-LEONI, M. E., CERCENADO, E., BOUZA, E., FENOLL, A. & SPRATT, B. G. 1996. Multiply antibiotic-resistant Streptococcus pneumoniae recovered from Spanish hospitals (1988-1994): novel major clones of serotypes 14, 19F and 15F. *Microbiology,* 142 (Pt 10)**,** 2747-57.
COFFEY, T. J., DANIELS, M., MCDOUGAL, L. K., DOWSON, C. G., TENOVER, F. C. & SPRATT, B. G. 1995. Genetic analysis of clinical isolates of Streptococcus pneumoniae with high-level resistance to expanded-spectrum cephalosporins. *Antimicrob Agents Chemother,* 39**,** 1306-13.
COFFEY, T. J., DOWSON, C. G., DANIELS, M., ZHOU, J., MARTIN, C., SPRATT, B. G. & MUSSER, J. M. 1991. Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthetic genes, in natural populations of Streptococcus pneumoniae. *Mol Microbiol,* 5**,** 2255-60.
COHEN, A. L., HARRISON, L. H., FARLEY, M. M., REINGOLD, A. L., HADLER, J., SCHAFFNER, W., LYNFIELD, R., THOMAS, A. R., CAMPSMITH, M., LI, J., SCHUCHAT, A. & MOORE, M. R. 2010. Prevention of invasive pneumococcal disease among HIV-infected adults in the era of childhood pneumococcal immunization. *Aids,* 24**,** 2253-62.
COHEN, R., BINGEN, E., VARON, E., DE LA ROCQUE, F., BRAHIMI, N., LEVY, C., BOUCHERAT, M., LANGUE, J. & GESLIN, P. 1997. Change in nasopharyngeal carriage of Streptococcus pneumoniae resulting from antibiotic therapy for acute otitis media in children. *Pediatr Infect Dis J,* 16**,** 555-60.
COLLART, F. R. & HUBERMAN, E. 1988. Cloning and sequence analysis of the human and Chinese hamster inosine-5'-monophosphate dehydrogenase cDNAs. *J Biol Chem,* 263**,** 15769-72.
COLLET, J. F. & MESSENS, J. 2010. Structure, function, and mechanism of thioredoxin proteins. *Antioxid Redox Signal,* 13**,** 1205-16.
COMPOUND INTEREST. 2014. *A Brief Overview of Classes of Antibiotics* [Online]. Available: http://www.compoundchem.com/2014/09/08/antibiotics/ [Accessed 6 July 2015].
COOKE, M. S., EVANS, M. D., DIZDAROGLU, M. & LUNEC, J. 2003. Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J,* 17**,** 1195-214.
COOPER, G., EDWARDS, M. & ROSENSTEIN, C. 1929. The separation of types among the pneumococci hitherto called group IV and the development of therapeutic antiserums for these types. *J Exp Med,* 49**,** 461-74.
CORANDER, J. & TANG, J. 2007. Bayesian analysis of population structure based on linked molecular information. *Math Biosci,* 205**,** 19-31.
CORANDER, J., WALDMANN, P. & SILLANPAA, M. J. 2003. Bayesian analysis of genetic differentiation between populations. *Genetics,* 163**,** 367-74.

CORNICK, J. E., EVERETT, D. B., BROUGHTON, C., DENIS, B. B., BANDA, D. L., CARROL, E. D. & PARRY, C. M. 2011. Invasive Streptococcus pneumoniae in Children, Malawi, 2004–2006. *Emerging Infectious Diseases,* 17**,** 1107-1109.

COSTELLO, A. 2015. *District burden and costs of severe pneumonia before and after introduction of pneumococcal vaccine in Malawi* [Online]. Available: http://www.ucl.ac.uk/igh/research/a-z/district-burden-and-costs..

COURTNEY, H. S. & LI, Y. 2013. Non-immune binding of human IgG to M-related proteins confers resistance to phagocytosis of group A streptococci in blood. *PLoS One,* 8**,** e78719.

CROUCHER, N. J., CHEWAPREECHA, C., HANAGE, W. P., HARRIS, S. R., MCGEE, L., VAN DER LINDEN, M., SONG, J. H., KO, K. S., DE LENCASTRE, H., TURNER, C., YANG, F., SA-LEAO, R., BEALL, B., KLUGMAN, K. P., PARKHILL, J., TURNER, P. & BENTLEY, S. D. 2014a. Evidence for soft selective sweeps in the evolution of pneumococcal multidrug resistance and vaccine escape. *Genome Biol Evol,* 6**,** 1589-602.

CROUCHER, N. J., FINKELSTEIN, J. A., PELTON, S. I., MITCHELL, P. K., LEE, G. M., PARKHILL, J., BENTLEY, S. D., HANAGE, W. P. & LIPSITCH, M. 2013. Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat Genet,* 45**,** 656-63.

CROUCHER, N. J., HANAGE, W. P., HARRIS, S. R., MCGEE, L., VAN DER LINDEN, M., DE LENCASTRE, H., SA-LEAO, R., SONG, J. H., KO, K. S., BEALL, B., KLUGMAN, K. P., PARKHILL, J., TOMASZ, A., KRISTINSSON, K. G. & BENTLEY, S. D. 2014b. Variable recombination dynamics during the emergence, transmission and 'disarming' of a multidrug-resistant pneumococcal clone. *BMC Biol,* 12**,** 49.

CROUCHER, N. J., HARRIS, S. R., BARQUIST, L., PARKHILL, J. & BENTLEY, S. D. 2012. A high-resolution view of genome-wide pneumococcal transformation. *PLoS Pathog,* 8**,** e1002745.

CROUCHER, N. J., HARRIS, S. R., FRASER, C., QUAIL, M. A., BURTON, J., VAN DER LINDEN, M., MCGEE, L., VON GOTTBERG, A., SONG, J. H., KO, K. S., PICHON, B., BAKER, S., PARRY, C. M., LAMBERTSEN, L. M., SHAHINAS, D., PILLAI, D. R., MITCHELL, T. J., DOUGAN, G., TOMASZ, A., KLUGMAN, K. P., PARKHILL, J., HANAGE, W. P. & BENTLEY, S. D. 2011. Rapid Pneumococcal Evolution in Response to Clinical Interventions. *Science,* 331**,** 430-434.

CROUCHER, N. J., PAGE, A. J., CONNOR, T. R., DELANEY, A. J., KEANE, J. A., BENTLEY, S. D., PARKHILL, J. & HARRIS, S. R. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res,* 43**,** e15.

CROUCHER, N. J., WALKER, D., ROMERO, P., LENNARD, N., PATERSON, G. K., BASON, N. C., MITCHELL, A. M., QUAIL, M. A., ANDREW, P. W., PARKHILL, J., BENTLEY, S. D. & MITCHELL, T. J. 2009. Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone Streptococcus pneumoniaeSpain23F ST81. *J Bacteriol,* 191**,** 1480-9.

CUNDELL, D. R., GERARD, N. P., GERARD, C., IDANPAAN-HEIKKILA, I. & TUOMANEN, E. I. 1995. Streptococcus pneumoniae anchor to activated human cells by the receptor for platelet-activating factor. *Nature,* 377**,** 435-8.

CYBULSKA, J., JELJASZEWICZ, J., LUND, E. & MUNKSGAARD, A. 1970. Prevalence of types of Diplococcus pneumoniae and their susceptibility to 30 antibiotics. *Chemotherapy,* 15**,** 304-16.
DABNEY, A., STOREY, J. D. & WARNES, G. 2004. Q-value estimation for false discovery rate control. *Medicine,* 344**,** 539-548.
DAGAN, R. & KLUGMAN, K. P. 2008. Impact of conjugate pneumococcal vaccines on antibiotic resistance. *Lancet Infect Dis,* 8**,** 785-95.
DAGAN, R., SHRIKER, O., HAZAN, I., LEIBOVITZ, E., GREENBERG, D., SCHLAEFFER, F. & LEVY, R. 1998. Prospective study to determine clinical relevance of detection of pneumococcal DNA in sera of children by PCR. *J Clin Microbiol,* 36**,** 669-73.
DAGKESSAMANSKAIA, A., MOSCOSO, M., HÉNARD, V., GUIRAL, S., OVERWEG, K., REUTER, M., MARTIN, B., WELLS, J. & CLAVERYS, J.-P. 2004. Interconnection of competence, stress and CiaR regulons in Streptococcus pneumoniae: competence triggers stationary phase autolysis of ciaR mutant cells. *Molecular Microbiology,* 51**,** 1071-1086.
DAUM, R. S., HOGERMAN, D., RENNELS, M. B., BEWLEY, K., MALINOSKI, F., ROTHSTEIN, E., REISINGER, K., BLOCK, S., KEYSERLING, H. & STEINHOFF, M. 1997. Infant Immunization with Pneumococcal CRM_{197} Vaccines: Effect of Saccharide Size on Immunogenicity and Interactions with Simultaneously Administered Vaccines. *The Journal of Infectious Diseases,* 176**,** 445-455.
DAVIES, T. A., HE, W., BUSH, K. & FLAMM, R. K. 2010. Affinity of ceftobiprole for penicillin-binding protein 2b in Streptococcus pneumoniae strains with various susceptibilities to penicillin. *Antimicrob Agents Chemother,* 54**,** 4510-2.
DE KRUIJFF, B., VAN DAM, V. & BREUKINK, E. 2008. Lipid II: a central component in bacterial cell wall synthesis and a target for antibiotics. *Prostaglandins Leukot Essent Fatty Acids,* 79**,** 117-21.
DE VEGA, M. 2013. The minimal Bacillus subtilis nonhomologous end joining repair machinery. *PLoS One,* 8**,** e64232.
DEL BECCARO, M. A., MENDELMAN, P. M., INGLIS, A. F., RICHARDSON, M. A., DUNCAN, N. O., CLAUSEN, C. R. & STULL, T. L. 1992. Bacteriology of acute otitis media: A new perspective. *The Journal of Pediatrics,* 120**,** 81-84.
DENAPAITE, D., CHI, F., MAURER, P., NOLTE, O. & HAKENBECK, R. 2007. Mechanisms of Penicillin 11 Resistance in Streptococcus pneumoniae-Targets, Gene Transfer, and Mutations. *Molecular Biology of Streptococci.* 1 ed. Norfolk: Horizon Bioscience.
DEPRISTO, M. A., BANKS, E., POPLIN, R., GARIMELLA, K. V., MAGUIRE, J. R., HARTL, C., PHILIPPAKIS, A. A., DEL ANGEL, G., RIVAS, M. A., HANNA, M., MCKENNA, A., FENNELL, T. J., KERNYTSKY, A. M., SIVACHENKO, A. Y., CIBULSKIS, K., GABRIEL, S. B., ALTSHULER, D. & DALY, M. J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet,* 43**,** 491-8.
DESAI, B. V. & MORRISON, D. A. 2007. Transformation in Streptococcus pneumoniae: formation of eclipse complex in a coiA mutant implicates CoiA in genetic recombination. *Molecular Microbiology,* 63**,** 1107-1117.
DESJARDINS, P. & CONKLIN, D. 2010. NanoDrop microvolume quantitation of nucleic acids. *J Vis Exp*.

DESSEN, A., MOUZ, N., GORDON, E., HOPKINS, J. & DIDEBERG, O. 2001. Crystal structure of PBP2x from a highly penicillin-resistant Streptococcus pneumoniae clinical isolate: a mosaic framework containing 83 mutations. *J Biol Chem,* 276**,** 45106-12.
DHARMADHIKARI, A. K., BHARAMBE, H., DHARMADHIKARI, J. A., D'SOUZA, J. S. & MATHUR, D. 2014. DNA damage by OH radicals produced using intense, ultrashort, long wavelength laser pulses. *Phys Rev Lett,* 112**,** 138105.
DI GUILMI, A. M., MOUZ, N., ANDRIEU, J.-P., HOSKINS, J., JASKUNAS, S. R., GAGNON, J., DIDEBERG, O. & VERNET, T. 1998. Identification, purification, and characterization of transpeptidase and glycosyltransferase domains of Streptococcus pneumoniae penicillin-binding protein 1a. *Journal of bacteriology,* 180**,** 5652-5659.
DIAS, R., FELIX, D., CANICA, M. & TROMBE, M. C. 2009. The highly conserved serine threonine kinase StkP of Streptococcus pneumoniae contributes to penicillin susceptibility independently from genes encoding penicillin-binding proteins. *BMC Microbiol,* 9**,** 121.
DIAZ, E., LOPEZ, R. & GARCIA, J. L. 1992. Role of the major pneumococcal autolysin in the atypical response of a clinical isolate of Streptococcus pneumoniae. *J Bacteriol,* 174**,** 5508-15.
DILLARD, J. P. & YOTHER, J. 1994. Genetic and molecular characterization of capsular polysaccharide biosynthesis in Streptococcus pneumoniae type 3. *Molecular Microbiology,* 12**,** 959-972.
DILLINGHAM, M. S. & KOWALCZYKOWSKI, S. C. 2008. RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol Mol Biol Rev,* 72**,** 642-71, Table of Contents.
DIXON, J. M., LIPINSKI, A. E. & GRAHAM, M. E. 1977. Detection and prevalence of pneumococci with increased resistance to penicillin. *Can Med Assoc J,* 117**,** 1159-61.
DIZDAROGLU, M., JARUGA, P., BIRINCIOGLU, M. & RODRIGUEZ, H. 2002. Free radical-induced damage to DNA: mechanisms and measurement. *Free Radic Biol Med,* 32**,** 1102-15.
DOCHEZ, A. R. & AVERY, O. T. 1915. Varieties of pneumococcus and their relation to lobar pneumonia. *J Exp Med,* 21**,** 114-32.
DOCHEZ, A. R. & AVERY, O. T. 1917. The elaboration of specific soluble substance by pneumococcus during growth. *J Exp Med,* 26**,** 477-93.
DOCHEZ, A. R. & GILLESPIE, L. J. 1913. A biologic classification of pneumococci by means of immunity reactions. *Journal of the American Medical Association,* 61**,** 727-732.
DONATI, C., HILLER, N. L., TETTELIN, H., MUZZI, A., CROUCHER, N. J., ANGIUOLI, S. V., OGGIONI, M., DUNNING HOTOPP, J. C., HU, F. Z., RILEY, D. R., COVACCI, A., MITCHELL, T. J., BENTLEY, S. D., KILIAN, M., EHRLICH, G. D., RAPPUOLI, R., MOXON, E. R. & MASIGNANI, V. 2010. Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species. *Genome Biol,* 11**,** R107.
DONG, Q. J., WANG, L. L., TIAN, Z. B., YU, X. J., JIA, S. J. & XUAN, S. Y. 2014. Reduced genome size of Helicobacter pylori originating from East Asia. *World J Gastroenterol,* 20**,** 5666-71.
DOWDS, B. C. & HOCH, J. A. 1991. Regulation of the oxidative stress response by the hpr gene in Bacillus subtilis. *J Gen Microbiol,* 137**,** 1121-5.

DOWSON, C. G., COFFEY, T. J., KELL, C. & WHILEY, R. A. 1993. Evolution of penicillin resistance in Streptococcus pneumoniae; the role of Streptococcus mitis in the formation of a low affinity PBP2B in S. pneumoniae. *Mol Microbiol,* 9**,** 635-43.

DOWSON, C. G., HUTCHISON, A., BRANNIGAN, J. A., GEORGE, R. C., HANSMAN, D., LINARES, J., TOMASZ, A., SMITH, J. M. & SPRATT, B. G. 1989. Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of Streptococcus pneumoniae. *Proc Natl Acad Sci U S A,* 86**,** 8842-6.

DOWSON, C. G., HUTCHISON, A., WOODFORD, N., JOHNSON, A. P., GEORGE, R. C. & SPRATT, B. G. 1990. Penicillin-resistant viridans streptococci have obtained altered penicillin-binding protein genes from penicillin-resistant strains of Streptococcus pneumoniae. *Proceedings of the National Academy of Sciences,* 87**,** 5858-5862.

DRASKOVIC, I. & DUBNAU, D. 2005. Biogenesis of a putative channel protein, ComEC, required for DNA uptake: membrane topology, oligomerization and formation of disulphide bonds. *Mol Microbiol,* 55**,** 881-96.

DROGE, W. 2002. Free radicals in the physiological control of cell function. *Physiol Rev,* 82**,** 47-95.

DU PLESSIS, M., SMITH, A. M. & KLUGMAN, K. P. 1998. Rapid detection of penicillin-resistant Streptococcus pneumoniae in cerebrospinal fluid by a seminested-PCR strategy. *J Clin Microbiol,* 36**,** 453-7.

DU PLESSIS, M., SMITH, A. M. & KLUGMAN, K. P. 1999. Application of pbp1A PCR in identification of penicillin-resistant Streptococcus pneumoniae. *J Clin Microbiol,* 37**,** 628-32.

DUANE, P. G., RUBINS, J. B., WEISEL, H. R. & JANOFF, E. N. 1993. Identification of hydrogen peroxide as a Streptococcus pneumoniae toxin for rat alveolar epithelial cells. *Infect Immun,* 61**,** 4392-7.

DUBOS, R. & AVERY, O. T. 1931. Decomposition of the capsular polysaccharide of pneumococcus type III by a bacterial enzyme. *J Exp Med,* 54**,** 51-71.

DUBOS, R. J. 1939a. Studies on a bactericidal agent extracted from a soil bacillus: I. Preparation of the agent. Its activity in vitro. *J Exp Med,* 70**,** 1-10.

DUBOS, R. J. 1939b. Studies on a bactericidal agent extracted from a soil bacillus: II. Protective effect of the bactericidal agent against experimental pneumococcus infections in mice. *J Exp Med,* 70**,** 11-7.

DUGGAR, B. M. 1948. Aureomycin: A product of the continuing search for new antibiotics. *Annals of the New York Academy of Sciences,* 51**,** 177-181.

DURAND, M. L., CALDERWOOD, S. B., WEBER, D. J., MILLER, S. I., SOUTHWICK, F. S., CAVINESS, V. S., JR. & SWARTZ, M. N. 1993. Acute bacterial meningitis in adults. A review of 493 episodes. *N Engl J Med,* 328**,** 21-8.

EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res,* 32**,** 1792-7.

EL KAROUI, M., BIAUDET, V., SCHBATH, S. & GRUSS, A. 1999. Characteristics of Chi distribution on different bacterial genomes. *Res Microbiol,* 150**,** 579-87.

EL ZOEIBY, A., SANSCHAGRIN, F. & LEVESQUE, R. C. 2003. Structure and function of the Mur enzymes: development of novel inhibitors. *Mol Microbiol,* 47**,** 1-12.

ENA. 2015. *European Nucleotide Archive* [Online]. Available: http://www.ebi.ac.uk/ena [Accessed 21 July 2015].
ENRIGHT, M. C. & SPRATT, B. G. 1998. A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. *Microbiology,* 144 (Pt 11)**,** 3049-60.
ENRIGHT, M. C. & SPRATT, B. G. 1999. Extensive variation in the ddl gene of penicillin-resistant Streptococcus pneumoniae results from a hitchhiking effect driven by the penicillin-binding protein 2b gene. *Mol Biol Evol,* 16**,** 1687-95.
ERIKSEN, K. R. 1945. Studies on induced resistance to penicillin in a pneumococcus of type I*. *Acta Pathologica Microbiologica Scandinavica,* 22**,** 398-405.
ERILL, I., CAMPOY, S. & BARBE, J. 2007. Aeons of distress: an evolutionary perspective on the bacterial SOS response. *FEMS Microbiol Rev,* 31**,** 637-56.
ERLICH, Y., MITRA, P. P., DELABASTIDE, M., MCCOMBIE, W. R. & HANNON, G. J. 2008. Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. *Nat Meth,* 5**,** 679-682.
EUCAST. 2015. *European Committee on Antimicrobial Susceptiblity Testing* [Online]. Available: http://www.eucast.org/ [Accessed 7 July 2015].
EVANS, G. M. & GAISFORD, W. F. 1938. Treatment of pnemonia with 2-(*p*-aminobenzenesulphonamido) pyridine. *Lancet,* 2**,** 14-19.
EVANS, M. D. & COOKE, M. S. 2004. Factors contributing to the outcome of oxidative damage to nucleic acids. *Bioessays,* 26**,** 533-42.
EVANS, W. & HANSMAN, D. 1963. Tetracycline-resistant pneumococcus. *Lancet,* 1**,** 451-&.
EVERETT, D. B., CORNICK, J., DENIS, B., CHEWAPREECHA, C., CROUCHER, N., HARRIS, S., PARKHILL, J., GORDON, S., CARROL, E. D., FRENCH, N., HEYDERMAN, R. S. & BENTLEY, S. D. 2012. Genetic characterisation of Malawian pneumococci prior to the roll-out of the PCV13 vaccine using a high-throughput whole genome sequencing approach. *PLoS One,* 7**,** e44250.
EVERETT, D. B., MUKAKA, M., DENIS, B., GORDON, S. B., CARROL, E. D., VAN OOSTERHOUT, J. J., MOLYNEUX, E. M., MOLYNEUX, M., FRENCH, N. & HEYDERMAN, R. S. 2011. Ten years of surveillance for invasive Streptococcus pneumoniae during the era of antiretroviral scale-up and cotrimoxazole prophylaxis in Malawi. *PLoS One,* 6**,** e17765.
EZRATY, B., BOS, J., BARRAS, F. & AUSSEL, L. 2005. Methionine sulfoxide reduction and assimilation in Escherichia coli: new role for the biotin sulfoxide reductase BisC. *J Bacteriol,* 187**,** 231-7.
FACKLAM, R. R. W., J. A 1991. *Streptococcus* and related catalase-negative gram-positive cocci. *Manual of Clinical Microbiology.* 5th ed. Washington American Society for Microbiology.
FADEN, H., DUFFY, L., WASIELEWSKI, R., WOLF, J., KRYSTOFIK, D. & TUNG, Y. 1997. Relationship between nasopharyngeal colonization and the development of otitis media in children. Tonawanda/Williamsville Pediatrics. *J Infect Dis,* 175**,** 1440-5.
FANG, G. D., FINE, M., ORLOFF, J., ARISUMI, D., YU, V. L., KAPOOR, W., GRAYSTON, J. T., WANG, S. P., KOHLER, R., MUDER, R. R. & ET AL. 1990. New and

emerging etiologies for community-acquired pneumonia with implications for therapy. A prospective multicenter study of 359 cases. *Medicine (Baltimore),* 69**,** 307-16.
FANI, F., LEPROHON, P., LEGARE, D. & OUELLETTE, M. 2011. Whole genome sequencing of penicillin-resistant Streptococcus pneumoniae reveals mutations in penicillin-binding proteins and in a putative iron permease. *Genome Biol,* 12**,** R115.
FARBER, B. F., ELIOPOULOS, G. M., WARD, J. I., RUOFF, K. L., SYRIOPOULOU, V. & MOELLERING, R. C., JR. 1983. Multiply resistant viridans streptococci: susceptibility to beta-lactam antibiotics and comparison of penicillin-binding protein patterns. *Antimicrob Agents Chemother,* 24**,** 702-5.
FARR, W. 1885. Vital Statistics: a memorial volume of selections from the report and writings of William Farr. 1 ed. London: Sanitary Institute.
FEDSON, D. S., NICOLAS-SPONY, L., KLEMETS, P., VAN DER LINDEN, M., MARQUES, A., SALLERAS, L. & SAMSON, S. I. 2011. Pneumococcal polysaccharide vaccination for adults: new perspectives for Europe. *Expert Rev Vaccines,* 10**,** 1143-67.
FEIKIN, D. R., DAVIS, M., NWANYANWU, O. C., KAZEMBE, P. N., BARAT, L. M., WASAS, A., BLOLAND, P. B., ZIBA, C., CAPPER, T., HUEBNER, R. E., SCHWARTZ, B., KLUGMAN, K. P. & DOWELL, S. F. 2003. Antibiotic resistance and serotype distribution of Streptococcus pneumoniae colonizing rural Malawian children. *Pediatr Infect Dis J,* 22**,** 564-7.
FEIL, E. J., COOPER, J. E., GRUNDMANN, H., ROBINSON, D. A., ENRIGHT, M. C., BERENDT, T., PEACOCK, S. J., SMITH, J. M., MURPHY, M., SPRATT, B. G., MOORE, C. E. & DAY, N. P. 2003. How clonal is Staphylococcus aureus? *J Bacteriol,* 185**,** 3307-16.
FEIL, E. J., SMITH, J. M., ENRIGHT, M. C. & SPRATT, B. G. 2000. Estimating recombinational parameters in Streptococcus pneumoniae from multilocus sequence typing data. *Genetics,* 154**,** 1439-50.
FELTON, L. D. 1924. A Study of the Isolation and Concentration of the Specific Antibodies of Antipneumococcus Sera. *The Boston Medical and Surgical Journal,* 190**,** 819-825.
FELTON, L. D., JORDAN, C. F., HESBACHER, E. N. & VAUBEL, E. K. 1941. Studies on immunizing substances in pneumococci: XII. Comparison of the effect of whole-cell vaccine and of polysaccharide antigen in human beings. *Public Health Reports,* 56**,** 1041-1054.
FERNANDEZ DE HENESTROSA, A. R., OGI, T., AOYAGI, S., CHAFIN, D., HAYES, J. J., OHMORI, H. & WOODGATE, R. 2000. Identification of additional genes belonging to the LexA regulon in Escherichia coli. *Mol Microbiol,* 35**,** 1560-72.
FERREIRA, D. M., NEILL, D. R., BANGERT, M., GRITZFELD, J. F., GREEN, N., WRIGHT, A. K., PENNINGTON, S. H., BRICIO-MORENO, L., MORENO, A. T., MIYAJI, E. N., WRIGHT, A. D., COLLINS, A. M., GOLDBLATT, D., KADIOGLU, A. & GORDON, S. B. 2013. Controlled human infection and rechallenge with Streptococcus pneumoniae reveals the protective efficacy of carriage in healthy adults. *Am J Respir Crit Care Med,* 187**,** 855-64.
FINKEL, T. & HOLBROOK, N. J. 2000. Oxidants, oxidative stress and the biology of ageing. *Nature,* 408**,** 239-47.

FLEMING, A. 1929. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. *British Journal of Experimental Pathology,* 10.
FRAENKEL, A. 1984. Uber die genuine Pneumonie. . *Verhandlugen des Congress fur innere Medicin.,* 3.
FRANCIS, T. & TILLETT, W. S. 1930. Cutaneous reactions in pneumonia. The development of antibodies following the intradermal injection of type-specific polysaccharide. *J Exp Med,* 52**,** 573-85.
FRANCO, R., SCHONEVELD, O., GEORGAKILAS, A. G. & PANAYIOTIDIS, M. I. 2008. Oxidative stress, DNA methylation and carcinogenesis. *Cancer Lett,* 266**,** 6-11.
FRENCH, N., GORDON, S. B., MWALUKOMO, T., WHITE, S. A., MWAFULIRWA, G., LONGWE, H., MWAIPONYA, M., ZIJLSTRA, E. E., MOLYNEUX, M. E. & GILKS, C. F. 2010. A trial of a 7-valent pneumococcal conjugate vaccine in HIV-infected adults. *N Engl J Med,* 362**,** 812-22.
FRIEDBERG, E. W., GC; SIEDE, W; WOOD, RD; SCHULTZ, RA; ELLENBERGER, T; 2006. *DNA repair and mutagenesis,* Washington DC, ASM Press.
FRISCH, A. W., PRICE, A. E. & MYERS, G. B. 1943. Type VIII pneumococcus: Development of sulfadiazine resistance, transmission by cross infection, and persistence in carriers. *Annals of Internal Medicine,* 18**,** 271-278.
FUJIMORI, I., KIKUSHIMA, K., HISAMATSU, K. I., NOZAWA, I., GOTO, R. & MURAKAMI, Y. 1996. Analysis of defense mechanisms against bacterial infection by oral streptococcus in normal flora. *Zentralbl Bakteriol,* 285**,** 74-81.
GAGNE, A. L., STEVENS, K. E., CASSONE, M., PUJARI, A., ABIOLA, O. E., CHANG, D. J. & SEBERT, M. E. 2013. Competence in Streptococcus pneumoniae Is a Response to an Increasing Mutational Burden. *PLoS ONE,* 8**,** e72613.
GARCIA, E., LLULL, D., MUNOZ, R., MOLLERACH, M. & LOPEZ, R. 2000. Current trends in capsular polysaccharide biosynthesis of Streptococcus pneumoniae. *Res Microbiol,* 151**,** 429-35.
GAVI 2013. Pneumococcal Disease. *In:* ALLIANCE, G. (ed.).
GAVI. 2015. *Pneumococcal Vaccine Support* [Online]. Available: http://www.gavi.org/support/nvs/pneumococcal/ [Accessed 6 July 2015].
GE, M., CHEN, Z., ONISHI, H. R., KOHLER, J., SILVER, L. L., KERNS, R., FUKUZAWA, S., THOMPSON, C. & KAHNE, D. 1999. Vancomycin derivatives that inhibit peptidoglycan biosynthesis without binding D-Ala-D-Ala. *Science,* 284**,** 507-11.
GEELEN, S., BHATTACHARYYA, C. & TUOMANEN, E. 1993. The cell wall mediates pneumococcal attachment to and cytopathology in human endothelial cells. *Infect Immun,* 61**,** 1538-43.
GENNARIS, A. & COLLET, J. F. 2013. The 'captain of the men of death', Streptococcus pneumoniae, fights oxidative stress outside the 'city wall'. *EMBO Mol Med,* 5**,** 1798-800.
GEORGE, R. C., COOPER, P. G. & ERDMAN, Y. J. 1987. Not the first multiresistant pneumococcus in Britain. *British Medical Journal (Clinical research ed.),* 295**,** 1208-1208.
GESSNER, B. D., MUELLER, J. E. & YARO, S. 2010. African meningitis belt pneumococcal disease epidemiology indicates a need for an effective

serotype 1 containing vaccine, including for older children and adults. *BMC Infect Dis,* 10**,** 22.
GHAFFAR, F., FRIEDLAND, I. R. & MCCRACKEN, G. H., JR. 1999. Dynamics of nasopharyngeal colonization by Streptococcus pneumoniae. *Pediatr Infect Dis J,* 18**,** 638-46.
GIBSON, L. F. & KHOURY, J. T. 1986. Storage and survival of bacteria by ultra-freeze. *Letters in Applied Microbiology,* 3**,** 127-129.
GIFFORD, M. L., BANTA, J. A., KATARI, M. S., HULSMANS, J., CHEN, L., RISTOVA, D., TRANCHINA, D., PURUGGANAN, M. D., CORUZZI, G. M. & BIRNBAUM, K. D. 2013. Plasticity Regulators Modulate Specific Root Traits in Discrete Nitrogen Environments. *PLoS Genet,* 9.
GIZI, A., PAPASSOTIRIOU, I., APOSTOLAKOU, F., LAZAROPOULOU, C., PAPASTAMATAKI, M., KANAVAKI, I., KALOTYCHOU, V., GOUSSETIS, E., KATTAMIS, A., ROMBOS, I. & KANAVAKIS, E. 2011. Assessment of oxidative stress in patients with sickle cell disease: The glutathione system and the oxidant-antioxidant status. *Blood Cells Mol Dis,* 46**,** 220-5.
GLADSTONE, R. A., JEFFERIES, J. M., TOCHEVA, A. S., BEARD, K. R., GARLEY, D., CHONG, W. W., BENTLEY, S. D., FAUST, S. N. & CLARKE, S. C. 2015. Five winters of pneumococcal serotype replacement in UK carriage following PCV introduction. *Vaccine,* 33**,** 2015-21.
GOLDENBERGER, D., KUNZLI, A., VOGT, P., ZBINDEN, R. & ALTWEGG, M. 1997. Molecular diagnosis of bacterial endocarditis by broad-range PCR amplification and direct sequencing. *J Clin Microbiol,* 35**,** 2733-9.
GORDON, M. A., WALSH, A. L., CHAPONDA, M., SOKO, D., MBVWINJI, M., MOLYNEUX, M. E. & GORDON, S. B. 2001. Bacteraemia and mortality among adult medical admissions in Malawi--predominance of non-typhi salmonellae and Streptococcus pneumoniae. *J Infect,* 42**,** 44-9.
GORDON, S. B., CHAPONDA, M., WALSH, A. L., WHITTY, C. J., GORDON, M. A., MACHILI, C. E., GILKS, C. F., BOEREE, M. J., KAMPONDENI, S., READ, R. C. & MOLYNEUX, M. E. 2002. Pneumococcal disease in HIV-infected Malawian adults: acute mortality and long-term survival. *AIDS,* 16**,** 1409-17.
GORDON, S. B., WALSH, A. L., CHAPONDA, M., GORDON, M. A., SOKO, D., MBWVINJI, M., MOLYNEUX, M. E. & READ, R. C. 2000. Bacterial Meningitis in Malawian Adults: Pneumococcal Disease is Common, Severe, and Seasonal. *Clinical Infectious Diseases,* 31**,** 53-57.
GOTTLIEB, D., ROBBINS, P. W. & CARTER, H. E. 1956. The biosynthesis of chloramphenicol: II. Acetylation ofp-Nitrophenylserinol. *J Bacteriol,* 72**,** 153-6.
GOULD, C. V., SNIEGOWSKI, P. D., SHCHEPETOV, M., METLAY, J. P. & WEISER, J. N. 2007. Identifying Mutator Phenotypes among Fluoroquinolone-Resistant Strains of Streptococcus pneumoniae Using Fluctuation Analysis. *Antimicrobial Agents and Chemotherapy,* 51**,** 3225-3229.
GRABENSTEIN, J. D. & KLUGMAN, K. P. 2012. A century of pneumococcal vaccination research in humans. *Clinicl Microbiology Infection,* 18 Suppl 5**,** 15-24.
GRAD, Y. H., KIRKCALDY, R. D., TREES, D., DORDEL, J., HARRIS, S. R., GOLDSTEIN, E., WEINSTOCK, H., PARKHILL, J., HANAGE, W. P., BENTLEY, S. & LIPSITCH, M. 2014. Genomic epidemiology of Neisseria gonorrhoeae with

reduced susceptibility to cefixime in the USA: a retrospective observational study. *Lancet Infect Dis,* 14**,** 220-6.
GRAM, C. 1884. Ueber die isolirte Farbung der Schizomyceten in Schnitt-und Trockenpraparaten. *Fortschritte der Medicin,* 2.
GRANGER, D., BOILY-LAROUCHE, G., TURGEON, P., WEISS, K. & ROGER, M. 2005. Genetic analysis of pbp2x in clinical Streptococcus pneumoniae isolates in Quebec, Canada. *J Antimicrob Chemother,* 55**,** 832-9.
GRANOFF, D. M., ANDERSON, E. L., OSTERHOLM, M. T., HOLMES, S. J., MCHUGH, J. E., BELSHE, R. B., MEDLEY, F. & MURPHY, T. V. 1992. Differences in the immunogenicity of three Haemophilus influenzae type b conjugate vaccines in infants. *The Journal of Pediatrics,* 121**,** 187-194.
GRATTEN, M., NARAQI, S. & HANSMAN, D. 1980. High prevalence of penicillin-insensitive pneumococci in Port Moresby, Papua New Guinea. *The Lancet,* 316**,** 192-195.
GRAY, B. M., CONVERSE, G. M., 3RD & DILLON, H. C., JR. 1980. Epidemiologic studies of Streptococcus pneumoniae in infants: acquisition, carriage, and infection during the first 24 months of life. *J Infect Dis,* 142**,** 923-33.
GRAY, B. M., TURNER, M. E. & DILLON, H. C., JR. 1982. Epidemiologic studies of Streptococcus pneumoniae in infants. The effects of season and age on pneumococcal acquisition and carriage in the first 24 months of life. *Am J Epidemiol,* 116**,** 692-703.
GREBE, T. & HAKENBECK, R. 1996. Penicillin-binding proteins 2b and 2x of Streptococcus pneumoniae are primary resistance determinants for different classes of beta-lactam antibiotics. *Antimicrob Agents Chemother,* 40**,** 829-34.
GREINERT, R., VOLKMER, B., HENNING, S., BREITBART, E. W., GREULICH, K. O., CARDOSO, M. C. & RAPP, A. 2012. UVA-induced DNA double-strand breaks result from the repair of clustered oxidative DNA damages. *Nucleic Acids Res,* 40**,** 10263-73.
GRIFFITH, F. 1928. The Significance of Pneumococcal Types. *J Hyg (Lond),* 27**,** 113-59.
GUENZI, E., GASC, A. M., SICARD, M. A. & HAKENBECK, R. 1994. A two-component signal-transducing system is involved in competence and penicillin susceptibility in laboratory mutants of Streptococcus pneumoniae. *Mol Microbiol,* 12**,** 505-15.
GUIRAL, S., HÉNARD, V., GRANADEL, C., MARTIN, B. & CLAVERYS, J.-P. 2006. Inhibition of competence development in Streptococcus pneumoniae by increased basal-level expression of the ComDE two-component regulatory system. *Microbiology,* 152**,** 323-331.
GUIRAL, S., MITCHELL, T. J., MARTIN, B. & CLAVERYS, J. P. 2005. Competence-programmed predation of noncompetent cells in the human pathogen Streptococcus pneumoniae: genetic requirements. *Proc Natl Acad Sci U S A,* 102**,** 8710-5.
HACKEL, M., LASCOLS, C., BOUCHILLON, S., HILTON, B., MORGENSTERN, D. & PURDY, J. 2013. Serotype prevalence and antibiotic resistance in Streptococcus pneumoniae clinical isolates among global populations. *Vaccine,* 31**,** 4881-7.

HAKENBECK, R., BRUCKNER, R., DENAPAITE, D. & MAURER, P. 2012. Molecular mechanisms of beta-lactam resistance in Streptococcus pneumoniae. *Future Microbiol,* 7**,** 395-410.
HAKENBECK, R., GREBE, T., ZAHNER, D. & STOCK, J. B. 1999. beta-lactam resistance in Streptococcus pneumoniae: penicillin-binding proteins and non-penicillin-binding proteins. *Mol Microbiol,* 33**,** 673-8.
HAKENBECK, R., KONIG, A., KERN, I., VAN DER LINDEN, M., KECK, W., BILLOT-KLEIN, D., LEGRAND, R., SCHOOT, B. & GUTMANN, L. 1998. Acquisition of five high-Mr penicillin-binding protein variants during transfer of high-level beta-lactam resistance from Streptococcus mitis to Streptococcus pneumoniae. *J Bacteriol,* 180**,** 1831-40.
HAKENBECK, R., TARPAY, M. & TOMASZ, A. 1980. Multiple changes of penicillin-binding proteins in penicillin-resistant clinical isolates of Streptococcus pneumoniae. *Antimicrob Agents Chemother,* 17**,** 364-71.
HALLIWELL, B. & ARUOMA, O. I. 1991. DNA damage by oxygen-derived species. Its mechanism and measurement in mammalian systems. *FEBS Lett,* 281**,** 9-19.
HAMBURGER, M., JR, SCHMIDT, L. H., RUEGSEGGER, J. M., SESLER, C. L. & GRUPEN, E. S. 1942. SUlfonamide resistance developing during treatment of pneumococcic endocarditis. *Journal of the American Medical Association,* 119**,** 409-411.
HAMMERSCHMIDT, S., WOLFF, S., HOCKE, A., ROSSEAU, S., MULLER, E. & ROHDE, M. 2005. Illustration of pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells. *Infect Immun,* 73**,** 4653-67.
HANAGE, W. P., BISHOP, C. J., HUANG, S. S., STEVENSON, A. E., PELTON, S. I., LIPSITCH, M. & FINKELSTEIN, J. A. 2011. Carried Pneumococci in Massachusetts Children; The Contribution of Clonal Expansion and Serotype Switching. *Pediatr Infect Dis J,* 30**,** 302-8.
HANAGE, W. P., FRASER, C. & SPRATT, B. G. 2006. Sequences, sequence clusters and bacterial species. *Philos Trans R Soc Lond B Biol Sci,* 361**,** 1917-27.
HANAGE, W. P., FRASER, C., TANG, J., CONNOR, T. R. & CORANDER, J. 2009. Hyper-recombination, diversity, and antibiotic resistance in pneumococcus. *Science,* 324**,** 1454-7.
HANSMAN, D. & BULLEN, M. M. 1967. A resistant pneumococcus. *Lancet,* 2**,** 264-&.
HARABUCHI, Y., FADEN, H., YAMANAKA, N., DUFFY, L., WOLF, J. & KRYSTOFIK, D. 1994. Nasopharyngeal colonization with nontypeable Haemophilus influenzae and recurrent otitis media. Tonawanda/Williamsville Pediatrics. *J Infect Dis,* 170**,** 862-6.
HARRIS, S. R., CARTWRIGHT, E. J., TOROK, M. E., HOLDEN, M. T., BROWN, N. M., OGILVY-STUART, A. L., ELLINGTON, M. J., QUAIL, M. A., BENTLEY, S. D., PARKHILL, J. & PEACOCK, S. J. 2013a. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. *Lancet Infect Dis,* 13**,** 130-6.
HARRIS, S. R., CLARKE, I. N., SETH-SMITH, H. M., SOLOMON, A. W., CUTCLIFFE, L. T., MARSH, P., SKILTON, R. J., HOLLAND, M. J., MABEY, D., PEELING, R. W., LEWIS, D. A., SPRATT, B. G., UNEMO, M., PERSSON, K., BJARTLING, C., BRUNHAM, R., DE VRIES, H. J., MORRE, S. A., SPEKSNIJDER, A., BEBEAR, C. M., CLERC, M., DE BARBEYRAC, B., PARKHILL, J. & THOMSON, N. R. 2012.

Whole-genome analysis of diverse Chlamydia trachomatis strains identifies phylogenetic relationships masked by current clinical typing. *Nat Genet,* 44**,** 413-9, S1.
HARRIS, S. R., ROBINSON, C., STEWARD, K. F., WEBB, K. S., PAILLOT, R., PARKHILL, J., HOLDEN, M. T. & WALLER, A. S. 2015. Genome specialization and decay of the strangles pathogen, Streptococcus equi, is driven by persistent infection. *Genome Res*.
HARRIS, S. R., TOROK, M. E., CARTWRIGHT, E. J., QUAIL, M. A., PEACOCK, S. J. & PARKHILL, J. 2013b. Read and assembly metrics inconsequential for clinical utility of whole-genome sequencing in mapping outbreaks. *Nat Biotechnol,* 31**,** 592-4.
HART, C. A. 1998. Antibiotic resistance: an increasing problem?. It always has been, but there are things we can do. *BMJ,* 316**,** 1255-6.
HART, M., STEEL, A., CLARK, S. A., MOYLE, G., NELSON, M., HENDERSON, D. C., WILSON, R., GOTCH, F., GAZZARD, B. & KELLEHER, P. 2007. Loss of discrete memory B cell subsets is associated with impaired immunization responses in HIV-1 infection and may be a risk factor for invasive pneumococcal disease. *J Immunol,* 178**,** 8212-20.
HASSAN-KING, M., BALDEH, I., SECKA, O., FALADE, A. & GREENWOOD, B. 1994. Detection of Streptococcus pneumoniae DNA in blood cultures by PCR. *J Clin Microbiol,* 32**,** 1721-4.
HAVA, D. L. & CAMILLI, A. 2002. Large-scale identification of serotype 4 Streptococcus pneumoniae virulence factors. *Mol Microbiol,* 45**,** 1389-406.
HÅVARSTEIN, L. S., HAKENBECK, R. & GAUSTAD, P. 1997. Natural competence in the genus Streptococcus: evidence that streptococci can change pherotype by interspecies recombinational exchanges. *J Bacteriol,* 179**,** 6589-94.
HAVARSTEIN, L. S., MARTIN, B., JOHNSBORG, O., GRANADEL, C. & CLAVERYS, J. P. 2006. New insights into the pneumococcal fratricide: relationship to clumping and identification of a novel immunity factor. *Mol Microbiol,* 59**,** 1297-307.
HEFFRON, R. 1939. Pneumonia with special reference to pneumococcus lobar pneumonia. *Journal of the American Medical Association,* 113**,** 2175-2175.
HEIDELBERGER, M. & AVERY, O. T. 1923. The soluble specific substance of pneumococcus. *J Exp Med,* 38**,** 73-9.
HEIDELBERGER, M. & HOBBY, G. L. 1942. Oxidized Cotton, an Immunologically Specific Polysaccharide. *Proc Natl Acad Sci U S A,* 28**,** 516-8.
HENRIQUES-NORMARK, B. & TUOMANEN, E. I. 2013. The pneumococcus: epidemiology, microbiology, and pathogenesis. *Cold Spring Harb Perspect Med,* 3.
HICKS, L. A., HARRISON, L. H., FLANNERY, B., HADLER, J. L., SCHAFFNER, W., CRAIG, A. S., JACKSON, D., THOMAS, A., BEALL, B., LYNFIELD, R., REINGOLD, A., FARLEY, M. M. & WHITNEY, C. G. 2007. Incidence of pneumococcal disease due to non-pneumococcal conjugate vaccine (PCV7) serotypes in the United States during the era of widespread PCV7 vaccination, 1998-2004. *J Infect Dis,* 196**,** 1346-54.
HILL, P. C., CHEUNG, Y. B., AKISANYA, A., SANKAREH, K., LAHAI, G., GREENWOOD, B. M. & ADEGBOLA, R. A. 2008. Nasopharyngeal carriage of

Streptococcus pneumoniae in Gambian infants: a longitudinal study. *Clin Infect Dis,* 46**,** 807-14.
HILLER, N. L., AHMED, A., POWELL, E., MARTIN, D. P., EUTSEY, R., EARL, J., JANTO, B., BOISSY, R. J., HOGG, J., BARBADORA, K., SAMPATH, R., LONERGAN, S., POST, J. C., HU, F. Z. & EHRLICH, G. D. 2010. Generation of Genic Diversity among <italic>Streptococcus pneumoniae</italic> Strains via Horizontal Gene Transfer during a Chronic Polyclonal Pediatric Infection. *PLoS Pathog,* 6**,** e1001108.
HIRST, R. A., KADIOGLU, A., O'CALLAGHAN, C. & ANDREW, P. W. 2004. The role of pneumolysin in pneumococcal pneumonia and meningitis. *Clin Exp Immunol,* 138**,** 195-201.
HJALMARSDOTTIR, M. A. & KRISTINSSON, K. G. 2014. Epidemiology of penicillin-non-susceptible pneumococci in Iceland, 1995-2010. *J Antimicrob Chemother,* 69**,** 940-6.
HO, P. L., WONG, R. C., CHOW, F. K., CHEUNG, M. Y., WONG, S. S., YAM, W. C. & QUE, T. L. 2004. Application of a multiplex pbp2b and pbp2x PCR for prediction of penicillin resistance in Streptococcus pneumoniae. *J Antimicrob Chemother,* 53**,** 890-1.
HODGSON, A. E., NELSON, S. M., BROWN, M. R. & GILBERT, P. 1995. A simple in vitro model for growth control of bacterial biofilms. *J Appl Bacteriol,* 79**,** 87-93.
HOLLINGSHEAD, S. K., BECKER, R. & BRILES, D. E. 2000. Diversity of PspA: mosaic genes and evidence for past recombination in Streptococcus pneumoniae. *Infect Immun,* 68**,** 5889-900.
HOLM, S. E. & GRAHN, E. 1983. Bacterial interference in streptococcal tonsillitis. *Scand J Infect Dis Suppl,* 39**,** 73-8.
HOLZAPFEL, W. H. & WOOD, B. J. B. 1995. Lactic acid bacteria in contemporary perspective. *In:* WOOD, B. J. B. & HOLZAPFEL, W. H. (eds.) *The Genera of Lactic Acid Bacteria.* Springer US.
HORSBURGH, M. J., CLEMENTS, M. O., CROSSLEY, H., INGHAM, E. & FOSTER, S. J. 2001. PerR controls oxidative stress resistance and iron storage proteins and is required for virulence in Staphylococcus aureus. *Infect Immun,* 69**,** 3744-54.
HOSKINS, J., ALBORN, W. E., JR., ARNOLD, J., BLASZCZAK, L. C., BURGETT, S., DEHOFF, B. S., ESTREM, S. T., FRITZ, L., FU, D. J., FULLER, W., GERINGER, C., GILMOUR, R., GLASS, J. S., KHOJA, H., KRAFT, A. R., LAGACE, R. E., LEBLANC, D. J., LEE, L. N., LEFKOWITZ, E. J., LU, J., MATSUSHIMA, P., MCAHREN, S. M., MCHENNEY, M., MCLEASTER, K., MUNDY, C. W., NICAS, T. I., NORRIS, F. H., O'GARA, M., PEERY, R. B., ROBERTSON, G. T., ROCKEY, P., SUN, P. M., WINKLER, M. E., YANG, Y., YOUNG-BELLIDO, M., ZHAO, G., ZOOK, C. A., BALTZ, R. H., JASKUNAS, S. R., ROSTECK, P. R., JR., SKATRUD, P. L. & GLASS, J. I. 2001. Genome of the bacterium Streptococcus pneumoniae strain R6. *J Bacteriol,* 183**,** 5709-17.
HOSKINS, J. A., MATSUSHIMA, P., MULLEN, D. L., TANG, J., ZHAO, G., MEIER, T. I., NICAS, T. I. & JASKUNAS, S. R. 1999. Gene Disruption Studies of Penicillin-Binding Proteins 1a, 1b, and 2a in Streptococcus pneumoniae. *J Bacteriol,* 181**,** 6552-5.

HOTCHKISS, R. D. & DUBOS, R. J. 1940. Fractionation of the bactericidal agent from cultures of soil bacillus. *Journal of Biological Chemistry,* 132**,** 791-792.
HOTCHKISS, R. D. & GOEBEL, W. F. 1937. Chemo-immunological studies on the soluble specific substance of pneumococcus: III. The structure of the aldobionic acid from the type III polysaccharide. *Journal of Biological Chemistry,* 121**,** 195-203.
HUI, F. M., ZHOU, L. & MORRISON, D. A. 1995. Competence for genetic transformation in Streptococcus pneumoniae: organization of a regulatory locus with homology to two lactococcin A secretion genes. *Gene,* 153**,** 25-31.
HUMBERT, O., PRUDHOMME, M., HAKENBECK, R., DOWSON, C. G. & CLAVERYS, J. P. 1995. Homeologous recombination and mismatch repair during transformation in Streptococcus pneumoniae: saturation of the Hex mismatch repair system. *Proc Natl Acad Sci U S A,* 92**,** 9052-6.
HUSSAIN, M., MELEGARO, A., PEBODY, R. G., GEORGE, R., EDMUNDS, W. J., TALUKDAR, R., MARTIN, S. A., EFSTRATIOU, A. & MILLER, E. 2005. A longitudinal household study of Streptococcus pneumoniae nasopharyngeal carriage in a UK setting. *Epidemiol Infect,* 133**,** 891-8.
IACCARINO, M. & BERG, P. 1971. Isoleucine auxotrophy as a consequence of a mutationally altered isoleucyl-transfer ribonucleic acid synthetase. *J Bacteriol,* 105**,** 527-37.
IMLAY, J. A. 2013. The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nat Rev Micro,* 11**,** 443-454.
INAMINE, G. S. & DUBNAU, D. 1995. ComEA, a Bacillus subtilis integral membrane protein required for genetic transformation, is needed for both DNA binding and transport. *J Bacteriol,* 177**,** 3045-51.
ISOZUMI, R., ITO, Y., ISHIDA, T., HIRAI, T., ITO, I., MANIWA, K., HAYASHI, M., KAGIOKA, H., HIRABAYASHI, M., ONARU, K., TOMIOKA, H., TOMII, K., GOHMA, I., OSAWA, M., IMAI, S., TAKAKURA, S., IINUMA, Y., CHIN, K., ICHIYAMA, S. & MISHIMA, M. 2008. Molecular characteristics of serotype 3 Streptococcus pneumoniae isolates among community-acquired pneumonia patients in Japan. *J Infect Chemother,* 14**,** 258-61.
IYER, R. R., PLUCIENNIK, A., BURDETT, V. & MODRICH, P. L. 2006. DNA mismatch repair: functions and mechanisms. *Chem Rev,* 106**,** 302-23.
JABES, D., NACHMAN, S. & TOMASZ, A. 1989. Penicillin-Binding Protein Families: Evidence for the Clonal Nature of Penicillin Resistance in Clinical Isolates of Pneumococci. *Journal of Infectious Diseases,* 159**,** 16-25.
JACOBS, M. R., KOORNHOF, H. J., ROBINS-BROWNE, R. M., STEVENSON, C. M., VERMAAK, Z. A., FREIMAN, I., MILLER, G. B., WITCOMB, M. A., ISAACSON, M., WARD, J. I. & AUSTRIAN, R. 1978. Emergence of multiply resistant pneumococci. *N Engl J Med,* 299**,** 735-40.
JAIN, V., KUMAR, M. & CHATTERJI, D. 2006. ppGpp: stringent response and survival. *J Microbiol,* 44**,** 1-10.
JAMES, D. B. & YOTHER, J. 2012. Genetic and biochemical characterizations of enzymes involved in Streptococcus pneumoniae serotype 2 capsule synthesis demonstrate that Cps2T (WchF) catalyzes the committed step

by addition of beta1-4 rhamnose, the second sugar residue in the repeat unit. *J Bacteriol,* 194**,** 6479-89.
JANDA, J. M. & ABBOTT, S. L. 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol,* 45**,** 2761-4.
JANOFF, E. N., BREIMAN, R. F., DALEY, C. L. & HOPEWELL, P. C. 1992. Pneumococcal disease during HIV infection. Epidemiologic, clinical, and immunologic perspectives. *Ann Intern Med,* 117**,** 314-24.
JANOFF, E. N., O'BRIEN, J., THOMPSON, P., EHRET, J., MEIKLEJOHN, G., DUVALL, G. & DOUGLAS, J. M., JR. 1993. Streptococcus pneumoniae colonization, bacteremia, and immune response among persons with human immunodeficiency virus infection. *J Infect Dis,* 167**,** 49-56.
JANULCZYK, R., IANNELLI, F., SJOHOLM, A. G., POZZI, G. & BJORCK, L. 2000. Hic, a novel surface protein of Streptococcus pneumoniae that interferes with complement function. *J Biol Chem,* 275**,** 37257-63.
JARVA, H., JANULCZYK, R., HELLWAGE, J., ZIPFEL, P. F., BJORCK, L. & MERI, S. 2002. Streptococcus pneumoniae evades complement attack and opsonophagocytosis by expressing the pspC locus-encoded Hic protein that binds to short consensus repeats 8-11 of factor H. *J Immunol,* 168**,** 1886-94.
JAUNEIKAITE, E., JEFFERIES, J. M., HIBBERD, M. L. & CLARKE, S. C. 2012. Prevalence of Streptococcus pneumoniae serotypes causing invasive and non-invasive disease in South East Asia: a review. *Vaccine,* 30**,** 3503-14.
JENSEN, A. & KILIAN, M. 2012. Delineation of Streptococcus dysgalactiae, its subspecies, and its clinical and phylogenetic relationship to Streptococcus pyogenes. *J Clin Microbiol,* 50**,** 113-26.
JOB, V., CARAPITO, R., VERNET, T., DESSEN, A. & ZAPUN, A. 2008. Common alterations in PBP1a from resistant Streptococcus pneumoniae decrease its reactivity toward beta-lactams: structural insights. *J Biol Chem,* 283**,** 4886-94.
JOHANSSON, N., KALIN, M., BACKMAN-JOHANSSON, C., LARSSON, A., NILSSON, K. & HEDLUND, J. 2014. Procalcitonin levels in community-acquired pneumonia - correlation with aetiology and severity. *Scand J Infect Dis,* 46**,** 787-91.
JOHNSTON, C., HINDS, J., SMITH, A., VAN DER LINDEN, M., VAN ELDERE, J. & MITCHELL, T. J. 2010. Detection of large numbers of pneumococcal virulence genes in streptococci of the mitis group. *J Clin Microbiol,* 48**,** 2762-9.
JOHNSTON, C., MARTIN, B., GRANADEL, C., POLARD, P. & CLAVERYS, J. P. 2013. Programmed protection of foreign DNA from restriction allows pathogenicity island exchange during pneumococcal transformation. *PLoS Pathog,* 9**,** e1003178.
JONES, N., HUEBNER, R., KHOOSAL, M., CREWE-BROWN, H. & KLUGMAN, K. 1998. The impact of HIV on Streptococcus pneumoniae bacteraemia in a South African population. *Aids,* 12**,** 2177-84.
JORGENSEN, J. H. & FERRARO, M. J. 2009. Antimicrobial susceptibility testing: a review of general principles and contemporary practices. *Clin Infect Dis,* 49**,** 1749-55.

KADIOGLU, A., WEISER, J. N., PATON, J. C. & ANDREW, P. W. 2008. The role of Streptococcus pneumoniae virulence factors in host respiratory colonization and disease. *Nat Rev Microbiol,* 6**,** 288-301.
KAMNG'ONA, A. W., HINDS, J., BAR-ZEEV, N., GOULD, K. A., CHAGUZA, C., MSEFULA, C., CORNICK, J. E., KULOHOMA, B. W., GRAY, K., BENTLEY, S. D., FRENCH, N., HEYDERMAN, R. S. & EVERETT, D. B. 2015. High multiple carriage and emergence of Streptococcus pneumoniae vaccine serotype variants in Malawian children. *BMC Infect Dis,* 15**,** 234.
KAPLAN, S. L., MASON, E. O., JR., WALD, E. R., SCHUTZE, G. E., BRADLEY, J. S., TAN, T. Q., HOFFMAN, J. A., GIVNER, L. B., YOGEV, R. & BARSON, W. J. 2004. Decrease of invasive pneumococcal infections in children among 8 children's hospitals in the United States after the introduction of the 7-valent pneumococcal conjugate vaccine. *Pediatrics,* 113**,** 443-9.
KAUFMAN, P. 1947. Pneumonia in old age: Active immunization against pneumonia with pneumococcus polysaccharide; results of a six year study. *Archives of Internal Medicine,* 79**,** 518-531.
KAWAGUCHI, K., OKU, N., RIN, K., YAMANAKA, K. & OKADA, S. 1996. Dimethylarsenics reveal DNA damage induced by superoxide anion radicals. *Biol Pharm Bull,* 19**,** 551-3.
KAWAMURA, Y., HOU, X. G., SULTANA, F., MIURA, H. & EZAKI, T. 1995. Determination of 16S rRNA sequences of Streptococcus mitis and Streptococcus gordonii and phylogenetic relationships among members of the genus Streptococcus. *Int J Syst Bacteriol,* 45**,** 406-8.
KAWAMURA, Y., WHILEY, R. A., SHU, S. E., EZAKI, T. & HARDIE, J. M. 1999. Genetic approaches to the identification of the mitis group within the genus Streptococcus. *Microbiology,* 145 (Pt 9)**,** 2605-13.
KAYHTY, H., AHMAN, H., RONNBERG, P. R., TILLIKAINEN, R. & ESKOLA, J. 1995. Pneumococcal polysaccharide-meningococcal outer membrane protein complex conjugate vaccine is immunogenic in infants and children. *J Infect Dis,* 172**,** 1273-8.
KEEFER, C. S., BLAKE, F. G., MARSHALL, E., JR, LOCKWOOD, J. S., WOOD, W. & JR 1943. Penicillin in the treatment of infections: A report of 500 cases. *Journal of the American Medical Association,* 122**,** 1217-1224.
KELLER, L. E., THOMAS, J. C., LUO, X., NAHM, M. H., MCDANIEL, L. S. & ROBINSON, D. A. 2013. Draft Genome Sequences of Five Multilocus Sequence Types of Nonencapsulated Streptococcus pneumoniae. *Genome Announc,* 1.
KIKUCHI, K., ENARI, T., TOTSUKA, K. & SHIMIZU, K. 1995. Comparison of phenotypic characteristics, DNA-DNA hybridization results, and results with a commercial rapid biochemical and enzymatic reaction system for identification of viridans group streptococci. *J Clin Microbiol,* 33**,** 1215-22.
KILIAN, M., MESTECKY, J. & RUSSELL, M. W. 1988. Defense mechanisms involving Fc-dependent functions of immunoglobulin A and their subversion by bacterial immunoglobulin A proteases. *Microbiol Rev,* 52**,** 296-303.
KILIAN, M., MIKKELSEN, L. & HENRICHSEN, J. 1989. Taxonomic Study of Viridans Streptococci: Description of Streptococcus gordonii sp. nov. and Emended Descriptions of Streptococcus sanguis (White and Niven 1946), Streptococcus oralis (Bridge and Sneath 1982), and Streptococcus mitis

(Andrewes and Horder 1906). *International Journal of Systematic and Evolutionary Microbiology,* 39**,** 471-484.
KILIAN, M., RILEY, D. R., JENSEN, A., BRUGGEMANN, H. & TETTELIN, H. 2014. Parallel evolution of Streptococcus pneumoniae and Streptococcus mitis to pathogenic and mutualistic lifestyles. *MBio,* 5**,** e01490-14.
KIM, J. B., PORRECA, G. J., SONG, L., GREENWAY, S. C., GORHAM, J. M., CHURCH, G. M., SEIDMAN, C. E. & SEIDMAN, J. G. 2007. Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy. *Science,* 316**,** 1481-4.
KIMURA, M. 1983. *The Neutral Theory of Molecular Evolution,* Cambridge, Cambridge University Press.
KING, S. J., HIPPE, K. R. & WEISER, J. N. 2006. Deglycosylation of human glycoconjugates by the sequential activities of exoglycosidases expressed by Streptococcus pneumoniae. *Mol Microbiol,* 59**,** 961-74.
KISLAK, J. W., RAZAVI, L. M., DALY, A. K. & FINLAND, M. 1965. Susceptibility of pneumococci to nine antibiotics. *Am J Med Sci,* 250**,** 261-8.
KLEIN, J. O. 1994. Otitis media. *Clin Infect Dis,* 19**,** 823-33.
KLEIN, J. O. 1997. Role of nontypeable Haemophilus influenzae in pediatric respiratory tract infections. *Pediatr Infect Dis J,* 16**,** S5-8.
KLEIN, M., KOEDEL, U. & PFISTER, H. W. 2006. Oxidative stress in pneumococcal meningitis: a future target for adjunctive therapy? *Prog Neurobiol,* 80**,** 269-80.
KLEMPERER, G. 1892. Die Beziehungen verschiedener Bakteriengifte zur Immunisirung und Heilung. *Zeitschrift für klinische Medizin,* 20.
KLEMPERER, G. K. F. 1891a. Versuche liber Immunisirung und Heilung bei der Pneumokokkeninfection. *Berlin klinische Wochenschrift,* 28.
KLEMPERER, G. K. F. 1891b. Versuche liber Immunisirung und Heilung bei der Pneumokokkeninfection. *Berlin klinische Wochenschrift,* 28.
KLUGMAN, K. P., COFFEY, T. J., SMITH, A., WASAS, A., MEYERS, M. & SPRATT, B. G. 1994. Cluster of an erythromycin-resistant variant of the Spanish multiply resistant 23F clone of Streptococcus pneumoniae in South Africa. *Eur J Clin Microbiol Infect Dis,* 13**,** 171-4.
KOEDEL, U., SCHELD, W. M. & PFISTER, H. W. 2002. Pathogenesis and pathophysiology of pneumococcal meningitis. *Lancet Infect Dis,* 2**,** 721-36.
KOHANSKI, M. A., DWYER, D. J. & COLLINS, J. J. 2010. How antibiotics kill bacteria: from targets to networks. *Nat Rev Microbiol,* 8**,** 423-35.
KOHANSKI, M. A., DWYER, D. J., HAYETE, B., LAWRENCE, C. A. & COLLINS, J. J. 2007. A common mechanism of cellular death induced by bactericidal antibiotics. *Cell,* 130**,** 797-810.
KOLKMAN, M. A., VAN DER ZEIJST, B. A. & NUIJTEN, P. J. 1998. Diversity of capsular polysaccharide synthesis gene clusters in Streptococcus pneumoniae. *J Biochem,* 123**,** 937-45.
KONONEN, E., JOUSIMIES-SOMER, H., BRYK, A., KILP, T. & KILIAN, M. 2002. Establishment of streptococci in the upper respiratory tract: longitudinal changes in the mouth and nasopharynx up to 2 years of age. *J Med Microbiol,* 51**,** 723-30.
KOWALCZYKOWSKI, S. C., DIXON, D. A., EGGLESTON, A. K., LAUDER, S. D. & REHRAUER, W. M. 1994. Biochemistry of homologous recombination in Escherichia coli. *Microbiol Rev,* 58**,** 401-65.

KRAUSS, J. & HAKENBECK, R. 1997. A mutation in the D,D-carboxypeptidase penicillin-binding protein 3 of Streptococcus pneumoniae contributes to cefotaxime resistance of the laboratory mutant C604. *Antimicrob Agents Chemother,* 41**,** 936-42.

KRETH, J., VU, H., ZHANG, Y. & HERZBERG, M. C. 2009. Characterization of Hydrogen Peroxide-Induced DNA Release by Streptococcus sanguinis and Streptococcus gordonii. *J Bacteriol,* 191**,** 6281-91.

KRONE, C. L., TRZCIŃSKI, K., ZBOROWSKI, T., SANDERS, E. A. M. & BOGAERT, D. 2013. Impaired Innate Mucosal Immunity in Aged Mice Permits Prolonged Streptococcus pneumoniae Colonization. *Infect Immun,* 81**,** 4615-25.

KULKARNI, A. & JULIN, D. A. 2004. Specific inhibition of the E.coli RecBCD enzyme by Chi sequences in single-stranded oligodeoxyribonucleotides. *Nucleic Acids Res,* 32**,** 3672-82.

KYAW, M. H., LYNFIELD, R., SCHAFFNER, W., CRAIG, A. S., HADLER, J., REINGOLD, A., THOMAS, A. R., HARRISON, L. H., BENNETT, N. M., FARLEY, M. M., FACKLAM, R. R., JORGENSEN, J. H., BESSER, J., ZELL, E. R., SCHUCHAT, A. & WHITNEY, C. G. 2006. Effect of Introduction of the Pneumococcal Conjugate Vaccine on Drug-Resistant Streptococcus pneumoniae. *New England Journal of Medicine,* 354**,** 1455-1463.

LACKS, S. 1966. Integration efficiency and genetic recombination in pneumococcal transformation. *Genetics,* 53**,** 207-35.

LACKS, S., GREENBERG, B. & NEUBERGER, M. 1974. Role of a Deoxyribonuclease in the Genetic Transformation of Diplococcus pneumoniae. *Proc Natl Acad Sci U S A,* 71**,** 2305-9.

LAIBLE, G., SPRATT, B. G. & HAKENBECK, R. 1991. Interspecies recombinational events during the evolution of altered PBP 2x genes in penicillin-resistant clinical isolates of Streptococcus pneumoniae. *Mol Microbiol,* 5**,** 1993-2002.

LAMBERT, P. A. 2002. Cellular impermeability and uptake of biocides and antibiotics in Gram-positive bacteria and mycobacteria. *J Appl Microbiol,* 92 Suppl**,** 46s-54s.

LANCEFIELD, R. C. 1933. A Serological Differentiation of Human and Other Groups of Hemolytic Streptococci. *J Exp Med,* 57**,** 571-95.

LAND, A. D., TSUI, H. C., KOCAOGLU, O., VELLA, S. A., SHAW, S. L., KEEN, S. K., SHAM, L. T., CARLSON, E. E. & WINKLER, M. E. 2013. Requirement of essential Pbp2x and GpsB for septal ring closure in Streptococcus pneumoniae D39. *Mol Microbiol,* 90**,** 939-55.

LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol,* 10**,** R25.

LAPEYSSONNIE, L. 1963. La méningite cérébro-spinale en Afrique. *Bulletin of the World Health Organisation,* 28**,** 3-114.

LATORRE, C., JUNCOSA, T. & SANFELIU, I. 1985. Antibiotic resistance and serotypes of 100 Streptococcus pneumoniae strains isolated in a children's hospital in Barcelona, Spain. *Antimicrob Agents Chemother,* 28**,** 357-9.

LAURENCEAU, R., PÉHAU-ARNAUDET, G., BACONNAIS, S., GAULT, J., MALOSSE, C., DUJEANCOURT, A., CAMPO, N., CHAMOT-ROOKE, J., LE CAM, E.,

CLAVERYS, J.-P. & FRONZES, R. 2013. A Type IV Pilus Mediates DNA Binding during Natural Transformation in <italic>Streptococcus pneumoniae</italic>. *PLoS Pathog,* 9**,** e1003473.
LEAMON, J. H., LEE, W. L., TARTARO, K. R., LANZA, J. R., SARKIS, G. J., DEWINTER, A. D., BERKA, J., WEINER, M., ROTHBERG, J. M. & LOHMAN, K. L. 2003. A massively parallel PicoTiterPlate based platform for discrete picoliter-scale polymerase chain reactions. *Electrophoresis,* 24**,** 3769-77.
LECERCLE, D., CLOUET, A., AL-DABBAGH, B., CROUVOISIER, M., BOUHSS, A., GRAVIER-PELLETIER, C. & LE MERRER, Y. 2010. Bacterial transferase MraY inhibitors: synthesis and biological evaluation. *Bioorg Med Chem,* 18**,** 4560-9.
LEDERBERG, J. & GOTSCHLICH, E. C. 2005. A Path to Discovery: The Career of Maclyn McCarty. *PLoS Biol,* 3.
LEE, M. S. & MORRISON, D. A. 1999. Identification of a new regulator in Streptococcus pneumoniae linking quorum sensing to competence for genetic transformation. *J Bacteriol,* 181**,** 5004-16.
LEWIS, R., NATHAN, N., DIARRA, L., BELANGER, F. & PAQUET, C. 2001. Timely detection of meningococcal meningitis epidemics in Africa. *Lancet,* 358**,** 287-93.
LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.
LI, H. & DURBIN, R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics,* 26**,** 589-95.
LIFE TECHNOLOGIES. 2015. *Qubit Fluorometer vs. Competitors* [Online]. Available: https://www.lifetechnologies.com/fr/fr/home/life-science/laboratory-instruments/fluorometers/qubit/qubit-fluorometer/qubit-vs-competitors.html [Accessed 8 July 2015].
LIM, D. & STRYNADKA, N. C. 2002. Structural basis for the beta lactam resistance of PBP2a from methicillin-resistant Staphylococcus aureus. *Nat Struct Biol,* 9**,** 870-6.
LINDHOLM, L. & SARKKINEN, H. 2004. Direct identification of gram-positive cocci from routine blood cultures by using AccuProbe tests. *J Clin Microbiol,* 42**,** 5609-13.
LIPSITCH, M. 1999. Bacterial vaccines and serotype replacement: lessons from Haemophilus influenzae and prospects for Streptococcus pneumoniae. *Emerg Infect Dis,* 5**,** 336-45.
LIPSITCH, M., ABDULLAHI, O., D'AMOUR, A., XIE, W., WEINBERGER, D. M., TCHETGEN TCHETGEN, E. & SCOTT, J. A. 2012. Estimating rates of carriage acquisition and clearance and competitive ability for pneumococcal serotypes in Kenya with a Markov transition model. *Epidemiology,* 23**,** 510-9.
LIPSITCH, M. & O'HAGAN, J. J. 2007. Patterns of antigenic diversity and the mechanisms that maintain them. *Journal of The Royal Society Interface,* 4**,** 787-802.
LIPSITCH, M., WHITNEY, C. G., ZELL, E., KAIJALAINEN, T., DAGAN, R. & MALLEY, R. 2005. Are anticapsular antibodies the primary mechanism of protection against invasive pneumococcal disease? *PLoS Med,* 2**,** e15.

LISTER, F. S. 1913. Specific serological reactions with pneumococci from different sources. *Publications of the South African Institute for Medical Research,* 1**,** 1-14.
LISTER, F. S. 1916. *An experimental study of prophylactic inoculation against pneumococcal infection in the rabbit and in man,* Johannesburg, W.E. Hortor & Co., Ltd.
LLOYD, A. J., GILBEY, A. M., BLEWETT, A. M., DE PASCALE, G., EL ZOEIBY, A., LEVESQUE, R. C., CATHERWOOD, A. C., TOMASZ, A., BUGG, T. D., ROPER, D. I. & DOWSON, C. G. 2008. Characterization of tRNA-dependent peptide bond formation by MurM in the synthesis of Streptococcus pneumoniae peptidoglycan. *J Biol Chem,* 283**,** 6402-17.
LLULL, D., MUÑOZ, R., LÓPEZ, R. & GARCÍA, E. 1999. A Single Gene (tts) Located outside the cap Locus Directs the Formation of Streptococcus pneumoniae Type 37 Capsular Polysaccharide: Type 37 Pneumococci Are Natural, Genetically Binary Strains. *J Exp Med,* 190**,** 241-52.
LOMAN, N. J., MISRA, R. V., DALLMAN, T. J., CONSTANTINIDOU, C., GHARBIA, S. E., WAIN, J. & PALLEN, M. J. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotech,* 30**,** 434-439.
LORENZ, M. G. & WACKERNAGEL, W. 1994. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev,* 58**,** 563-602.
LUESINK, E. J., VAN HERPEN, R. E., GROSSIORD, B. P., KUIPERS, O. P. & DE VOS, W. M. 1998. Transcriptional activation of the glycolytic las operon and catabolite repression of the gal operon in Lactococcus lactis are mediated by the catabolite control protein CcpA. *Mol Microbiol,* 30**,** 789-98.
LUO, P., LI, H. & MORRISON, D. A. 2004. Identification of ComW as a new component in the regulation of genetic transformation in Streptococcus pneumoniae. *Mol Microbiol,* 54**,** 172-83.
LUO, P. & MORRISON, D. A. 2003. Transient Association of an Alternative Sigma Factor, ComX, with RNA Polymerase during the Period of Competence for Genetic Transformation in Streptococcus pneumoniae. *J Bacteriol,* 185**,** 349-58.
LUOTONEN, J. 1982. Streptococcus pneumoniae and Haemophilus influenzae in nasal cultures during acute otitis media. *Acta Otolaryngol,* 93**,** 295-9.
LYNCH, J. P., 3RD & ZHANEL, G. G. 2010. Streptococcus pneumoniae: epidemiology and risk factors, evolution of antimicrobial resistance, and impact of vaccines. *Curr Opin Pulm Med,* 16**,** 217-25.
MACHEBOEUF, P., CONTRERAS-MARTEL, C., JOB, V., DIDEBERG, O. & DESSEN, A. 2006. Penicillin binding proteins: key players in bacterial cell cycle and drug resistance processes. *FEMS Microbiol Rev,* 30**,** 673-91.
MACLEOD, C. M. & DADDI, G. 1939. A "Sulfapyridine-Fast" Strain of Pneumococcus Type I. *Experimental Biology and Medicine,* 41**,** 69-71.
MACLEOD, C. M., HODGES, R. G., HEIDELBERGER, M. & BERNHARD, W. G. 1945. Prevention of pneumococcal pneumonia by immunisation with specific capsular polysaccharides. *J Exp Med,* 82**,** 445-65.
MACLEOD, C. M., MIRICK, G. S. & CURNEN, E. C. 1940. Toxicity for Dogs of a Bactericidal Substance Derived from a Soil Bacillus. *Experimental Biology and Medicine,* 43**,** 461-463.
MADHI, S. A., ADRIAN, P., KUWANDA, L., CUTLAND, C., ALBRICH, W. C. & KLUGMAN, K. P. 2007. Long-term effect of pneumococcal conjugate

vaccine on nasopharyngeal colonization by Streptococcus pneumoniae--and associated interactions with Staphylococcus aureus and Haemophilus influenzae colonization--in HIV-Infected and HIV-uninfected children. *J Infect Dis,* 196**,** 1662-6.

MAIDEN, M. C., BYGRAVES, J. A., FEIL, E., MORELLI, G., RUSSELL, J. E., URWIN, R., ZHANG, Q., ZHOU, J., ZURTH, K., CAUGANT, D. A., FEAVERS, I. M., ACHTMAN, M. & SPRATT, B. G. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A,* 95**,** 3140-5.

MAKOMBE, S., LIBAMBA, E., MHANGO, E., DE ASCURRA TECK, O., ABERLE-GRASSE, J., HOCHGESANG, M., SCHOUTEN, E. J. & HARRIES, A. D. 2006. Who is accessing antiretroviral therapy during national scale-up in Malawi? *Trans R Soc Trop Med Hyg,* 100**,** 975-9.

MANDELL, L. A., PETERSON, L. R., WISE, R., HOOPER, D., LOW, D. E., SCHAAD, U. B., KLUGMAN, K. P. & COURVALIN, P. 2002. The battle against emerging antibiotic resistance: should fluoroquinolones be used to treat children? *Clin Infect Dis,* 35**,** 721-7.

MANN, J. M., CARABETTA, V. J., CRISTEA, I. M. & DUBNAU, D. 2013. Complex formation and processing of the minor transformation pilins of Bacillus subtilis. *Molecular Microbiology,* 90**,** 1201-1215.

MARGOLIS, E., YATES, A. & LEVIN, B. R. 2010. The ecology of nasal colonization of Streptococcus pneumoniae, Haemophilus influenzae and Staphylococcus aureus: the role of competition and interactions with host's immune response. *BMC Microbiol,* 10**,** 59.

MARGULIES, M., EGHOLM, M., ALTMAN, W. E., ATTIYA, S., BADER, J. S., BEMBEN, L. A., BERKA, J., BRAVERMAN, M. S., CHEN, Y.-J., CHEN, Z., DEWELL, S. B., DU, L., FIERRO, J. M., GOMES, X. V., GODWIN, B. C., HE, W., HELGESEN, S., HO, C. H., IRZYK, G. P., JANDO, S. C., ALENQUER, M. L. I., JARVIE, T. P., JIRAGE, K. B., KIM, J.-B., KNIGHT, J. R., LANZA, J. R., LEAMON, J. H., LEFKOWITZ, S. M., LEI, M., LI, J., LOHMAN, K. L., LU, H., MAKHIJANI, V. B., MCDADE, K. E., MCKENNA, M. P., MYERS, E. W., NICKERSON, E., NOBILE, J. R., PLANT, R., PUC, B. P., RONAN, M. T., ROTH, G. T., SARKIS, G. J., SIMONS, J. F., SIMPSON, J. W., SRINIVASAN, M., TARTARO, K. R., TOMASZ, A., VOGT, K. A., VOLKMER, G. A., WANG, S. H., WANG, Y., WEINER, M. P., YU, P., BEGLEY, R. F. & ROTHBERG, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437**,** 376-380.

MARKIEWICZ, Z. & TOMASZ, A. 1989. Variation in penicillin-binding protein patterns of penicillin-resistant clinical isolates of pneumococci. *J Clin Microbiol,* 27**,** 405-10.

MARSCHALL, T., COSTA, I. G., CANZAR, S., BAUER, M., KLAU, G. W., SCHLIEP, A. & SCHONHUTH, A. 2012. CLEVER: clique-enumerating variant finder. *Bioinformatics,* 28**,** 2875-82.

MARSCHALL, T., HAJIRASOULIHA, I. & SCHONHUTH, A. 2013. MATE-CLEVER: Mendelian-inheritance-aware discovery and genotyping of midsize and long indels. *Bioinformatics,* 29**,** 3143-50.

MARSCHALL, T. S., ALEXANDER. 2013. *Sensitive Long-Indel-Aware Alignment of Sequencing Reads* [Online]. Centrum Wiskunde & Informatica, Amsterdam, Netherlands. Available: http://arxiv.org/pdf/1303.3520v1.pdf [Accessed 25 July 2015].

MARTENS, P., WORM, S. W., LUNDGREN, B., KONRADSEN, H. B. & BENFIELD, T. 2004. Serotype-specific mortality from invasive Streptococcus pneumoniae disease revisited. *BMC Infect Dis,* 4**,** 21.
MARTIN, B., HUMBERT, O., CAMARA, M., GUENZI, E., WALKER, J., MITCHELL, T., ANDREW, P., PRUDHOMME, M., ALLOING, G. & HAKENBECK, R. 1992. A highly conserved repeated DNA element located in the chromosome of Streptococcus pneumoniae. *Nucleic Acids Res,* 20**,** 3479-83.
MARTIN, B., SOULET, A. L., MIROUZE, N., PRUDHOMME, M., MORTIER-BARRIERE, I., GRANADEL, C., NOIROT-GROS, M. F., NOIROT, P., POLARD, P. & CLAVERYS, J. P. 2013. ComE/ComE~P interplay dictates activation or extinction status of pneumococcal X-state (competence). *Mol Microbiol,* 87**,** 394-411.
MARTIN, J. A. & WANG, Z. 2011. Next-generation transcriptome assembly. *Nat Rev Genet,* 12**,** 671-82.
MARTIN-GALIANO, A. J., BALSALOBRE, L., FENOLL, A. & DE LA CAMPA, A. G. 2003. Genetic characterization of optochin-susceptible viridans group streptococci. *Antimicrob Agents Chemother,* 47**,** 3187-94.
MASCHER, T., HEINTZ, M., ZAHNER, D., MERAI, M. & HAKENBECK, R. 2006. The CiaRH system of Streptococcus pneumoniae prevents lysis during stress induced by treatment with cell wall inhibitors and by mutations in pbp2x involved in beta-lactam resistance. *J Bacteriol,* 188**,** 1959-68.
MAURER, P., KOCH, B., ZERFASS, I., KRAUSS, J., VAN DER LINDEN, M., FRERE, J. M., CONTRERAS-MARTEL, C. & HAKENBECK, R. 2008. Penicillin-binding protein 2x of Streptococcus pneumoniae: three new mutational pathways for remodelling an essential enzyme into a resistance determinant. *J Mol Biol,* 376**,** 1403-16.
MAXAM, A. M. & GILBERT, W. 1977. A new method for sequencing DNA. *Proc Natl Acad Sci U S A,* 74**,** 560-4.
MAY, J. R. 1954. Pathogenic bacteria in chronic bronchitis. *Lancet,* 267**,** 839-42.
MAYNARD, G. D. 1913. An enquiry into the etiology, manifestations, and prevention of pneumonia amongst natives on the Rand recruited from tropical areas. *Publications of the South African Institute for Medical Research,* 1**,** 1-101.
MCALLISTER, L. J., TSENG, H. J., OGUNNIYI, A. D., JENNINGS, M. P., MCEWAN, A. G. & PATON, J. C. 2004. Molecular analysis of the psa permease complex of Streptococcus pneumoniae. *Molecular microbiology,* 53**,** 889-901.
MCCULLERS, J. A. & REHG, J. E. 2002. Lethal synergism between influenza virus and Streptococcus pneumoniae: characterization of a mouse model and the role of platelet-activating factor receptor. *J Infect Dis,* 186**,** 341-50.
MCDANIEL, L. S., SHEFFIELD, J. S., DELUCCHI, P. & BRILES, D. E. 1991. PspA, a surface protein of Streptococcus pneumoniae, is capable of eliciting protection against pneumococci of more than one capsular type. *Infect Immun,* 59**,** 222-8.
MCDOUGAL, L. K., FACKLAM, R., REEVES, M., HUNTER, S., SWENSON, J. M., HILL, B. C. & TENOVER, F. C. 1992. Analysis of multiply antimicrobial-resistant isolates of Streptococcus pneumoniae from the United States. *Antimicrob Agents Chemother,* 36**,** 2176-84.
MCGEE, L., MCDOUGAL, L., ZHOU, J., SPRATT, B. G., TENOVER, F. C., GEORGE, R., HAKENBECK, R., HRYNIEWICZ, W., LEFEVRE, J. C., TOMASZ, A. &

KLUGMAN, K. P. 2001. Nomenclature of major antimicrobial-resistant clones of Streptococcus pneumoniae defined by the pneumococcal molecular epidemiology network. *J Clin Microbiol,* 39**,** 2565-71.
MCKEE, C. M. & HOUCK, C. L. 1943. Induced Resistance to Penicillin of Cultures of Staphylococci, Pneumococci and Streptococci. *Experimental Biology and Medicine,* 53**,** 33-34.
MCKENNA, A., HANNA, M., BANKS, E., SIVACHENKO, A., CIBULSKIS, K., KERNYTSKY, A., GARIMELLA, K., ALTSHULER, D., GABRIEL, S., DALY, M. & DEPRISTO, M. A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res,* 20**,** 1297-303.
MCNALLY, L. M., JEENA, P. M., GAJEE, K., STURM, A. W., TOMKINS, A. M., COOVADIA, H. M. & GOLDBLATT, D. 2006. Lack of Association between the Nasopharyngeal Carriage of Streptococcus pneumoniae and Staphylococcus aureus in HIV-1–Infected South African Children. *Journal of Infectious Diseases,* 194**,** 385-390.
MEDVEDEV, P., STANCIU, M. & BRUDNO, M. 2009. Computational methods for discovering structural variation with next-generation sequencing. *Nat Methods,* 6**,** S13-20.
MEJEAN, V. & CLAVERYS, J. P. 1984. Use of a cloned DNA fragment to analyze the fate of donor DNA in transformation of Streptococcus pneumoniae. *J Bacteriol,* 158**,** 1175-8.
MEJEAN, V. & CLAVERYS, J. P. 1988. Polarity of DNA entry in transformation of Streptococcus pneumoniae. *Mol Gen Genet,* 213**,** 444-8.
MELLROTH, P., DANIELS, R., EBERHARDT, A., RONNLUND, D., BLOM, H., WIDENGREN, J., NORMARK, S. & HENRIQUES-NORMARK, B. 2012. LytA, major autolysin of Streptococcus pneumoniae, requires access to nascent peptidoglycan. *J Biol Chem,* 287**,** 11018-29.
METCHNIKOFF, E. 1891. Études sur l'immunité, 4e memoire. L'immunité de cobayes vaccinés contre le vibrio Metchnikowii. *Annales de l'Institut Pasteur,* 5.
METZKER, M. L. 2010. Sequencing technologies - the next generation. *Nat Rev Genet,* 11**,** 31-46.
MIAO, R. & GUILD, W. R. 1970. Competent Diplococcus pneumoniae accept both single- and double-stranded deoxyribonucleic acid. *J Bacteriol,* 101**,** 361-4.
MILES, A. A., MISRA, S. S. & IRWIN, J. O. 1938. The estimation of the bactericidal power of the blood. *J Hyg (Lond),* 38**,** 732-49.
MILLER, E., ANDREWS, N. J., WAIGHT, P. A., SLACK, M. P. & GEORGE, R. C. 2011. Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: an observational cohort study. *Lancet Infect Dis,* 11**,** 760-8.
MILLER, M. L., GAO, G., PESTINA, T., PERSONS, D. & TUOMANEN, E. 2007. Hypersusceptibility to Invasive Pneumococcal Infection in Experimental Sickle Cell Disease Involves Platelet-Activating Factor Receptor. *Journal of Infectious Diseases,* 195**,** 581-584.
MILLER, M. T., BACHMANN, B. O., TOWNSEND, C. A. & ROSENZWEIG, A. C. 2001. Structure of beta-lactam synthetase reveals how to synthesize antibiotics instead of asparagine. *Nature Structural Biology,* 8**,** 684-689.

MOELLERING, R. C., JR. 2006. Vancomycin: a 50-year reassessment. *Clin Infect Dis,* 42 Suppl 1**,** S3-4.
MOJA, P., JALIL, A., QUESNEL, A., PEROL, M., COTTE, L., LIVROZET, J. M., BOIBIEUX, A., CHAMSON, A., VERGNON, J. M., LUCHT, F., TRAN, R., POZZETTO, B. & GENIN, C. 1997. Humoral immune response within the lung in HIV-1 infection. *Clin Exp Immunol,* 110**,** 341-8.
MOLESWORTH, A. M., THOMSON, M. C., CONNOR, S. J., CRESSWELL, M. P., MORSE, A. P., SHEARS, P., HART, C. A. & CUEVAS, L. E. 2002. Where is the meningitis belt? Defining an area at risk of epidemic meningitis in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 96**,** 242-249.
MOORE, H. F. 1915. The chemoserotherapy of experimental pneumococal infection. *J Exp Med,* 22**,** 389-400.
MOORE, H. F. & CHESNEY, A. M. 1917. A study of ethylhydrocuprein (optochin) in the treatment of acute lobar pneumonia. *Archives of Internal Medicine,* XIX**,** 611-682.
MOORE, M. R., GERTZ, R. E., JR., WOODBURY, R. L., BARKOCY-GALLAGHER, G. A., SCHAFFNER, W., LEXAU, C., GERSHMAN, K., REINGOLD, A., FARLEY, M., HARRISON, L. H., HADLER, J. L., BENNETT, N. M., THOMAS, A. R., MCGEE, L., PILISHVILI, T., BRUEGGEMANN, A. B., WHITNEY, C. G., JORGENSEN, J. H. & BEALL, B. 2008. Population snapshot of emergent Streptococcus pneumoniae serotype 19A in the United States, 2005. *J Infect Dis,* 197**,** 1016-27.
MORENS, D. M., TAUBENBERGER, J. K. & FAUCI, A. S. 2008. Predominant role of bacterial pneumonia as a cause of death in pandemic influenza: implications for pandemic influenza preparedness. *J Infect Dis,* 198**,** 962-70.
MORGENROTH, J. & LEVY, R. 1911. Chemotherapie der Pneumokokkeninfektion. *Berlin klinische Wochenschrift,* 48.
MORLOT, C., PERNOT, L., LE GOUELLEC, A., DI GUILMI, A. M., VERNET, T., DIDEBERG, O. & DESSEN, A. 2005. Crystal structure of a peptidoglycan synthesis regulatory factor (PBP3) from Streptococcus pneumoniae. *J Biol Chem,* 280**,** 15984-91.
MORRISON, D. A. & GUILD, W. R. 1972. Transformation and deoxyribonucleic acid size: extent of degradation on entry varies with size of donor. *J Bacteriol,* 112**,** 1157-68.
MORRISON, D. A. & GUILD, W. R. 1973. Breakage prior to entry of donor DNA in Pneumococcus transformation. *Biochim Biophys Acta,* 299**,** 545-56.
MORTIER-BARRIERE, I., VELTEN, M., DUPAIGNE, P., MIROUZE, N., PIETREMENT, O., MCGOVERN, S., FICHANT, G., MARTIN, B., NOIROT, P., LE CAM, E., POLARD, P. & CLAVERYS, J. P. 2007. A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. *Cell,* 130**,** 824-36.
MULLER, L. M., GORTER, K. J., HAK, E., GOUDZWAARD, W. L., SCHELLEVIS, F. G., HOEPELMAN, A. I. & RUTTEN, G. E. 2005. Increased risk of common infections in patients with type 1 and type 2 diabetes mellitus. *Clin Infect Dis,* 41**,** 281-8.
MUNOZ, R., COFFEY, T. J., DANIELS, M., DOWSON, C. G., LAIBLE, G., CASAL, J., HAKENBECK, R., JACOBS, M., MUSSER, J. M., SPRATT, B. G. & ET AL. 1991.

Intercontinental spread of a multiresistant clone of serotype 23F Streptococcus pneumoniae. *J Infect Dis,* 164**,** 302-6.
MUNOZ, R., MUSSER, J. M., CRAIN, M., BRILES, D. E., MARTON, A., PARKINSON, A. J., SORENSEN, U. & TOMASZ, A. 1992. Geographic distribution of penicillin-resistant clones of Streptococcus pneumoniae: characterization by penicillin-binding protein profile, surface protein A typing, and multilocus enzyme analysis. *Clin Infect Dis,* 15**,** 112-8.
MURDOCH, J. M. C., SPEIRS, C. F., GEDDES, A. M. & WALLACE, E. T. 1964. Clinical Trial of Cephaloridine (Ceporin), a New Broad-spectrum Antibiotic Derived from Cephalosporin C. *Br Med J,* 2**,** 1238-40.
MURPHY, S. J., CHEVILLE, J. C., ZAREI, S., JOHNSON, S. H., SIKKINK, R. A., KOSARI, F., FELDMAN, A. L., ECKLOFF, B. W., KARNES, R. J. & VASMATZIS, G. 2012. Mate pair sequencing of whole-genome-amplified DNA following laser capture microdissection of prostate cancer. *DNA Res,* 19**,** 395-406.
MUSCHIOL, S., BALABAN, M., NORMARK, S. & HENRIQUES-NORMARK, B. 2015. Uptake of extracellular DNA: competence induced pili in natural transformation of Streptococcus pneumoniae. *Bioessays,* 37**,** 426-35.
NAKAMURA, S., DAVIS, K. M. & WEISER, J. N. 2011. Synergistic stimulation of type I interferons during influenza virus coinfection promotes Streptococcus pneumoniae colonization in mice. *The Journal of Clinical Investigation,* 121**,** 3657-3665.
NARAQI, S., KIRKPATRICK, G. P. & KABINS, S. 1974. Relapsing pneumococcal meningitis: isolation of an organism with decreased susceptibility to penicillin G. *J Pediatr,* 85**,** 671-3.
NEELEMAN, C., GEELEN, S. P., AERTS, P. C., DAHA, M. R., MOLLNES, T. E., ROORD, J. J., POSTHUMA, G., VAN DIJK, H. & FLEER, A. 1999. Resistance to both complement activation and phagocytosis in type 3 pneumococci is mediated by the binding of complement regulatory protein factor H. *Infect Immun,* 67**,** 4517-24.
NEUFELD, F. 1900. Über eine spezifische bakteriolytische Wirkung der Galle. *Zeitschrift fur Hygiene und Infektionskrankheiten,* 34.
NEUFELD, F. & ETINGER-TULCZNSKA, R. 1930. Untersuchungen zur Gallenlosung der Pneumokokken. *Archiv für Hygiene,* 103**,** 107.
NEUFELD, F. & HAENDEL, L. 1910. Weitere Untersuchungen über Pneumokokken-Heilsera. *Arbeitein aus dem K. Gesundheitsamt,* 34.
NICOLA, G., PEDDI, S., STEFANOVA, M., NICHOLAS, R. A., GUTHEIL, W. G. & DAVIES, C. 2005. Crystal structure of Escherichia coli penicillin-binding protein 5 bound to a tripeptide boronic acid inhibitor: a role for Ser-110 in deacylation. *Biochemistry,* 44**,** 8207-17.
NIELSEN, S. V. & HENRICHSEN, J. 1992. Capsular types of Streptococcus pneumoniae isolated from blood and CSF during 1982-1987. *Clin Infect Dis,* 15**,** 794-8.
NOVAK, R., CHARPENTIER, E., BRAUN, J. S. & TUOMANEN, E. 2000. Signal Transduction by a Death Signal Peptide. *Molecular Cell,* 5**,** 49-57.
NOVAK, R., HENRIQUES, B., CHARPENTIER, E., NORMARK, S. & TUOMANEN, E. 1999. Emergence of vancomycin tolerance in Streptococcus pneumoniae. *Nature,* 399**,** 590-3.
NUNES, M. C., SHIRI, T., VAN NIEKERK, N., CUTLAND, C. L., GROOME, M. J., KOEN, A., VON GOTTBERG, A., DE GOUVEIA, L., KLUGMAN, K. P., ADRIAN, P. V. &

MADHI, S. A. 2013. Acquisition of Streptococcus pneumoniae in pneumococcal conjugate vaccine-naive South African children and their mothers. *Pediatr Infect Dis J,* 32**,** e192-205.
NUNES, S., SA-LEAO, R. & DE LENCASTRE, H. 2008. Optochin resistance among Streptococcus pneumoniae strains colonizing healthy children in Portugal. *J Clin Microbiol,* 46**,** 321-4.
NUORTI, J. P., BUTLER, J. C., FARLEY, M. M., HARRISON, L. H., MCGEER, A., KOLCZAK, M. S. & BREIMAN, R. F. 2000. Cigarette smoking and invasive pneumococcal disease. Active Bacterial Core Surveillance Team. *N Engl J Med,* 342**,** 681-9.
NYIRENDA, L. J., SANDBERG, K. I. & JUSTICE, J. 2014. When Are Health Systems Ready for New Vaccines? The Introduction of Pneumococcal Vaccine in Malawi. *Forum for Development Studies,* 41**,** 317-336.
O'BRIEN, K. L., BRONSDON, M. A., DAGAN, R., YAGUPSKY, P., JANCO, J., ELLIOTT, J., WHITNEY, C. G., YANG, Y. H., ROBINSON, L. G., SCHWARTZ, B. & CARLONE, G. M. 2001. Evaluation of a medium (STGG) for transport and optimal recovery of Streptococcus pneumoniae from nasopharyngeal secretions collected during field studies. *J Clin Microbiol,* 39**,** 1021-4.
O'BRIEN, K. L., NOHYNEK, H. & WORLD HEALTH ORGANIZATION PNEUMOCOCCAL VACCINE TRIALS CARRIAGE WORKING, G. 2003. Report from a WHO Working Group: standard method for detecting upper respiratory carriage of Streptococcus pneumoniae. *Pediatr Infect Dis J,* 22**,** e1-11.
O'BRIEN, K. L., WOLFSON, L. J., WATT, J. P., HENKLE, E., DELORIA-KNOLL, M., MCCALL, N., LEE, E., MULHOLLAND, K., LEVINE, O. S., CHERIAN, T., HIB & PNEUMOCOCCAL GLOBAL BURDEN OF DISEASE STUDY, T. 2009. Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. *Lancet,* 374**,** 893-902.
OBERT, C., SUBLETT, J., KAUSHAL, D., HINOJOSA, E., BARTON, T., TUOMANEN, E. I. & ORIHUELA, C. J. 2006. Identification of a Candidate Streptococcus pneumoniae core genome and regions of diversity correlated with invasive pneumococcal disease. *Infect Immun,* 74**,** 4766-77.
OBREGON, V., GARCIA, P., GARCIA, E., FENOLL, A., LOPEZ, R. & GARCIA, J. L. 2002. Molecular peculiarities of the lytA gene isolated from clinical pneumococcal strains that are bile insoluble. *J Clin Microbiol,* 40**,** 2545-54.
OGUNNIYI, A. D., MAHDI, L. K., TRAPPETTI, C., VERHOEVEN, N., MERMANS, D., VAN DER HOEK, M. B., PLUMPTRE, C. D. & PATON, J. C. 2012. Identification of Genes That Contribute to the Pathogenesis of Invasive Pneumococcal Disease by In Vivo Transcriptomic Analysis. *Infect Immun,* 80**,** 3268-78.
ORIO, A. G. A., PINAS, G. E., CORTES, P. R., CIAN, M. B. & ECHENIQUE, J. 2011. Compensatory evolution of pbp mutations restores the fitness cost imposed by beta-lactam resistance in Streptococcus pneumoniae. *PLoS Pathog,* 7**,** e1002000.
OTTO, T. D., DILLON, G. P., DEGRAVE, W. S. & BERRIMAN, M. 2011. RATT: Rapid Annotation Transfer Tool. *Nucleic Acids Res,* 39**,** e57.
PAGLIERO, E., CHESNEL, L., HOPKINS, J., CROIZE, J., DIDEBERG, O., VERNET, T. & DI GUILMI, A. M. 2004. Biochemical characterization of Streptococcus

pneumoniae penicillin-binding protein 2b and its implication in beta-lactam resistance. *Antimicrob Agents Chemother,* 48**,** 1848-55.
PAI, R., MOORE, M. R., PILISHVILI, T., GERTZ, R. E., WHITNEY, C. G. & BEALL, B. 2005. Postvaccine genetic structure of Streptococcus pneumoniae serotype 19A from children in the United States. *J Infect Dis,* 192**,** 1988-95.
PANE, N. 1897. Ueber die Heilkraft des aus verschiedenen immunisierten Tieren gewonnenen antipneumonischen Serums. *Centralblatt für Bakteriologie,* 21.
PARADISO, P. R. 2011. Advances in Pneumococcal Disease Prevention: 13-Valent Pneumococcal Conjugate Vaccine for Infants and Children. *Clinical Infectious Diseases,* 52**,** 1241-1247.
PARK, H. K., DANG, H. T., MYUNG, S. C. & KIM, W. 2012. Identification of a pheA gene associated with Streptococcus mitis by using suppression subtractive hybridization. *Appl Environ Microbiol,* 78**,** 3004-9.
PARK, H. K., LEE, H. J., JEONG, E. G., SHIN, H. S. & KIM, W. 2010. The rgg gene is a specific marker for Streptococcus oralis. *J Dent Res,* 89**,** 1299-303.
PARK, J. T. & UEHARA, T. 2008. How Bacteria Consume Their Own Exoskeletons (Turnover and Recycling of Cell Wall Peptidoglycan). *Microbiol Mol Biol Rev,* 72**,** 211-27.
PARRY, C. M., DUONG, N. M., ZHOU, J., MAI, N. T., DIEP, T. S., THINH LE, Q., WAIN, J., VAN VINH CHAU, N., GRIFFITHS, D., DAY, N. P., WHITE, N. J., HIEN, T. T., SPRATT, B. G. & FARRAR, J. J. 2002. Emergence in Vietnam of Streptococcus pneumoniae resistant to multiple antimicrobial agents as a result of dissemination of the multiresistant Spain(23F)-1 clone. *Antimicrob Agents Chemother,* 46**,** 3512-7.
PASTEUR, L., CHAMBERLAND, ROUX 1881. Sur une maladie nouvelle, provoquée par la salive d'un enfant mort de la rage. *Comptes Rendues de'lAcadémie des Sciences,* 92.
PASTOR, P., MEDLEY, F. & MURPHY, T. V. 1998. Invasive pneumococcal disease in Dallas County, Texas: results from population-based surveillance in 1995. *Clin Infect Dis,* 26**,** 590-5.
PATERSON, G. K., BLUE, C. E. & MITCHELL, T. J. 2006. An operon in Streptococcus pneumoniae containing a putative alkylhydroperoxidase D homologue contributes to virulence and the response to oxidative stress. *Microb Pathog,* 40**,** 152-60.
PATERSON, G. K. & MITCHELL, T. J. 2006. Innate immunity and the pneumococcus. *Microbiology,* 152**,** 285-93.
PATERSON, G. K., NIEMINEN, L., JEFFERIES, J. M. & MITCHELL, T. J. 2008. PclA, a pneumococcal collagen-like protein with selected strain distribution, contributes to adherence and invasion of host cells. *FEMS Microbiol Lett,* 285**,** 170-6.
PEREZ-NUNEZ, D., BRIANDET, R., DAVID, B., GAUTIER, C., RENAULT, P., HALLET, B., HOLS, P., CARBALLIDO-LOPEZ, R. & GUEDON, E. 2011. A new morphogenesis pathway in bacteria: unbalanced activity of cell wall synthesis machineries leads to coccus-to-rod transition and filamentation in ovococci. *Mol Microbiol,* 79**,** 759-71.
PERICONE, C. D., BAE, D., SHCHEPETOV, M., MCCOOL, T. & WEISER, J. N. 2002. Short-sequence tandem and nontandem DNA repeats and endogenous

hydrogen peroxide production contribute to genetic instability of Streptococcus pneumoniae. *J Bacteriol,* 184**,** 4392-9.
PERICONE, C. D., OVERWEG, K., HERMANS, P. W. & WEISER, J. N. 2000. Inhibitory and bactericidal effects of hydrogen peroxide production by Streptococcus pneumoniae on other inhabitants of the upper respiratory tract. *Infect Immun,* 68**,** 3990-7.
PERICONE, C. D., PARK, S., IMLAY, J. A. & WEISER, J. N. 2003. Factors contributing to hydrogen peroxide resistance in Streptococcus pneumoniae include pyruvate oxidase (SpxB) and avoidance of the toxic effects of the fenton reaction. *J Bacteriol,* 185**,** 6815-25.
PERNOT, L., CHESNEL, L., LE GOUELLEC, A., CROIZE, J., VERNET, T., DIDEBERG, O. & DESSEN, A. 2004. A PBP2x from a clinical isolate of Streptococcus pneumoniae exhibits an alternative mechanism for reduction of susceptibility to beta-lactam antibiotics. *J Biol Chem,* 279**,** 16463-70.
PESTOVA, E. V., HAVARSTEIN, L. S. & MORRISON, D. A. 1996. Regulation of competence for genetic transformation in Streptococcus pneumoniae by an auto-induced peptide pheromone and a two-component regulatory system. *Mol Microbiol,* 21**,** 853-62.
PETERSON, S. N., SUNG, C. K., CLINE, R., DESAI, B. V., SNESRUD, E. C., LUO, P., WALLING, J., LI, H., MINTZ, M., TSEGAYE, G., BURR, P. C., DO, Y., AHN, S., GILBERT, J., FLEISCHMANN, R. D. & MORRISON, D. A. 2004. Identification of competence pheromone responsive genes in Streptococcus pneumoniae by use of DNA microarrays. *Mol Microbiol,* 51**,** 1051-70.
PIKIS, A., CAMPOS, J. M., RODRIGUEZ, W. J. & KEITH, J. M. 2001. Optochin resistance in Streptococcus pneumoniae: mechanism, significance, and clinical implications. *J Infect Dis,* 184**,** 582-90.
PILISHVILI, T., LEXAU, C., FARLEY, M. M., HADLER, J., HARRISON, L. H., BENNETT, N. M., REINGOLD, A., THOMAS, A., SCHAFFNER, W., CRAIG, A. S., SMITH, P. J., BEALL, B. W., WHITNEY, C. G. & MOORE, M. R. 2010. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. *J Infect Dis,* 201**,** 32-41.
PIOTROWSKI, A., LUO, P. & MORRISON, D. A. 2009. Competence for genetic transformation in Streptococcus pneumoniae: termination of activity of the alternative sigma factor ComX is independent of proteolysis of ComX and ComW. *J Bacteriol,* 191**,** 3359-66.
PLOUFFE, J. F., BREIMAN, R. F. & FACKLAM, R. R. 1996. Bacteremia with Streptococcus pneumoniae. Implications for therapy and prevention. Franklin County Pneumonia Study Group. *JAMA,* 275**,** 194-8.
POEHLING, K. A., TALBOT, T. R., GRIFFIN, M. R., CRAIG, A. S., WHITNEY, C. G., ZELL, E., LEXAU, C. A., THOMAS, A. R., HARRISON, L. H., REINGOLD, A. L., HADLER, J. L., FARLEY, M. M., ANDERSON, B. J. & SCHAFFNER, W. 2006. Invasive pneumococcal disease among infants before and after introduction of pneumococcal conjugate vaccine. *Jama,* 295**,** 1668-74.
POLACK, F. P., FLAYHART, D. C., ZAHURAK, M. L., DICK, J. D. & WILLOUGHBY, R. E. 2000. Colonization by Streptococcus penumoniae in human immunodeficiency virus-infected children. *Pediatr Infect Dis J,* 19**,** 608-12.
POUTREL, B. & RYNIEWICZ, H. Z. 1984. Evaluation of the API 20 Strep system for species identification of streptococci isolated from bovine mastitis. *J Clin Microbiol,* 19**,** 213-4.

POWARS, D., OVERTURF, G., WEISS, J., LEE, S. & CHAN, L. 1981. Pneumococcal septicemia in children with sickle cell anemia. Changing trend of survival. *Jama,* 245**,** 1839-42.
POYART, C., QUESNE, G., COULON, S., BERCHE, P. & TRIEU-CUOT, P. 1998. Identification of streptococci to species level by sequencing the gene encoding the manganese-dependent superoxide dismutase. *J Clin Microbiol,* 36**,** 41-7.
POZZI, G., MASALA, L., IANNELLI, F., MANGANELLI, R., HAVARSTEIN, L. S., PICCOLI, L., SIMON, D. & MORRISON, D. A. 1996. Competence for genetic transformation in encapsulated strains of Streptococcus pneumoniae: two allelic variants of the peptide pheromone. *J Bacteriol,* 178**,** 6087-90.
PRADO, S., VILLARROYA, M., MEDINA, M. & ARMENGOD, M. E. 2013. The tRNA-modifying function of MnmE is controlled by post-hydrolysis steps of its GTPase cycle. *Nucleic Acids Res,* 41**,** 6190-208.
PREZA, D., OLSEN, I., AAS, J. A., WILLUMSEN, T., GRINDE, B. & PASTER, B. J. 2008. Bacterial profiles of root caries in elderly patients. *J Clin Microbiol,* 46**,** 2015-21.
PRO-LAB DIAGNOSTICS. 2012. *McFarland Standards* [Online]. Available: http://www.pro-lab.com/inserts/McFarland.pdf [Accessed 8 July 2015].
PROBER, J. M., TRAINOR, G. L., DAM, R. J., HOBBS, F. W., ROBERTSON, C. W., ZAGURSKY, R. J., COCUZZA, A. J., JENSEN, M. A. & BAUMEISTER, K. 1987. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science,* 238**,** 336-41.
PROVVEDI, R. & DUBNAU, D. 1999. ComEA is a DNA receptor for transformation of competent Bacillus subtilis. *Mol Microbiol,* 31**,** 271-80.
PROZOROV, A. A. & DANILENKO, V. N. 2011. Allolysis in bacteria. *Microbiology,* 80**,** 1-9.
PRUDHOMME, M., ATTAIECH, L., SANCHEZ, G., MARTIN, B. & CLAVERYS, J. P. 2006. Antibiotic stress induces genetic transformability in the human pathogen Streptococcus pneumoniae. *Science,* 313**,** 89-92.
PUBLIC HEALTH ENGLAND. 2014. *UK Standards for Microbiology and Investigations* [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/369824/ID_4i3.pdf [Accessed 9 July 2015].
PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M A R., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P I W., DALY, M J. & SHAM, P C. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet,* 81**,** 559-75.
PUYET, A., GREENBERG, B. & LACKS, S. A. 1990. Genetic and structural characterization of endA. A membrane-bound nuclease required for transformation of Streptococcus pneumoniae. *J Mol Biol,* 213**,** 727-38.
QUAIL, M. A., SMITH, M., COUPLAND, P., OTTO, T. D., HARRIS, S. R., CONNOR, T. R., BERTONI, A., SWERDLOW, H. P. & GU, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics,* 13**,** 341.
QUIBERONI, A., BISWAS, I., EL KAROUI, M., REZAIKI, L., TAILLIEZ, P. & GRUSS, A. 2001. In vivo evidence for two active nuclease motifs in the double-strand break repair enzyme RexAB of Lactococcus lactis. *J Bacteriol,* 183**,** 4071-8.

R, A. 2014. *Alpha and Beta Haemolysis* [Online]. Available: http://www.medicotips.com/2011/04/what-is-alpha-and-beta-hemolysis.html [Accessed 21 July 2015].
RAJAM, G., ANDERTON, J. M., CARLONE, G. M., SAMPSON, J. S. & ADES, E. W. 2008. Pneumococcal surface adhesin A (PsaA): a review. *Crit Rev Microbiol,* 34**,** 131-42.
RAMDANI-BOUGUESSA, N. & RAHAL, K. 2003. Serotype Distribution and Antimicrobial Resistance of Streptococcus pneumoniae Isolated in Algiers, Algeria. *Antimicrob Agents Chemother,* 47**,** 824-6.
READ, T. D. & MASSEY, R. C. 2014. Characterizing the genetic basis of bacterial phenotypes using genome-wide association studies: a new direction for bacteriology. *Genome Med,* 6**,** 109.
REDFIELD, R. J. 1993a. Evolution of Natural Transformation: Testing the DNA Repair Hypothesis in Bacillus Subtilis and Haemophilus Influenzae. *Genetics,* 133**,** 755-61.
REDFIELD, R. J. 1993b. Genes for breakfast: the have-your-cake-and-eat-it-too of bacterial transformation. *J Hered,* 84**,** 400-4.
REDMOND, R. W. & KOCHEVAR, I. E. 2006. Spatially resolved cellular responses to singlet oxygen. *Photochem Photobiol,* 82**,** 1178-86.
REGEV-YOCHAY, G., DAGAN, R., RAZ, M., CARMELI, Y., SHAINBERG, B., DERAZNE, E., RAHAV, G. & RUBINSTEIN, E. 2004. Association between carriage of Streptococcus pneumoniae and Staphylococcus aureus in Children. *Jama,* 292**,** 716-20.
REGEV-YOCHAY, G., TRZCINSKI, K., THOMPSON, C. M., LIPSITCH, M. & MALLEY, R. 2007. SpxB is a suicide gene of Streptococcus pneumoniae and confers a selective advantage in an in vivo competitive colonization model. *J Bacteriol,* 189**,** 6532-9.
REICHMANN, P., KONIG, A., LINARES, J., ALCAIDE, F., TENOVER, F. C., MCDOUGAL, L., SWIDSINSKI, S. & HAKENBECK, R. 1997. A global gene pool for high-level cephalosporin resistance in commensal Streptococcus species and Streptococcus pneumoniae. *J Infect Dis,* 176**,** 1001-12.
REINERT, R. R. 2009. The antimicrobial resistance profile of Streptococcus pneumoniae. *Clin Microbiol Infect,* 15 Suppl 3**,** 7-11.
REIS, J. N., PALMA, T., RIBEIRO, G. S., PINHEIRO, R. M., RIBEIRO, C. T., CORDEIRO, S. M., DA SILVA FILHO, H. P., MOSCHIONI, M., THOMPSON, T. A., SPRATT, B., RILEY, L. W., BAROCCHI, M. A., REIS, M. G. & KO, A. I. 2008. Transmission of Streptococcus pneumoniae in an urban slum community. *J Infect,* 57**,** 204-13.
REN, B., SZALAI, A. J., HOLLINGSHEAD, S. K. & BRILES, D. E. 2004. Effects of PspA and Antibodies to PspA on Activation and Deposition of Complement on the Pneumococcal Surface. *Infect Immun,* 72**,** 114-22.
REN, B., SZALAI, A. J., THOMAS, O., HOLLINGSHEAD, S. K. & BRILES, D. E. 2003. Both Family 1 and Family 2 PspA Proteins Can Inhibit Complement Deposition and Confer Virulence to a Capsular Serotype 3 Strain of Streptococcus pneumoniae. *Infect Immun,* 71**,** 75-85.
RENNELS, M. B., EDWARDS, K. M., KEYSERLING, H. L., REISINGER, K. S., HOGERMAN, D. A., MADORE, D. V., CHANG, I., PARADISO, P. R., MALINOSKI, F. J. & KIMURA, A. 1998. Safety and immunogenicity of

heptavalent pneumococcal vaccine conjugated to CRM197 in United States infants. *Pediatrics,* 101**,** 604-11.
REPINE, J. E., FOX, R. B. & BERGER, E. M. 1981. Hydrogen peroxide kills Staphylococcus aureus by reacting with staphylococcal iron to form hydroxyl radical. *J Biol Chem,* 256**,** 7094-6.
RILEY, I. D., LEHMANN, D. & ALPERS, M. P. 1991. Pneumococcal vaccine trials in Papua New Guinea: relationships between epidemiology of pneumococcal infection and efficacy of vaccine. *Rev Infect Dis,* 13 Suppl 6**,** S535-41.
RILEY, I. D., TARR, P. I., ANDREWS, M., PFEIFFER, M., HOWARD, R., CHALLANDS, P. & JENNISON, G. 1977. Immunisation with a polyvalent pneumococcal vaccine. Reduction of adult respiratory mortality in a New Guinea Highlands community. *Lancet,* 1**,** 1338-41.
ROBBINS, J. B. & SCHNEERSON, R. 1990. Polysaccharide-protein conjugates: a new generation of vaccines. *J Infect Dis,* 161**,** 821-32.
ROBINSON, K. A., BAUGHMAN, W., ROTHROCK, G., BARRETT, N. L., PASS, M., LEXAU, C., DAMASKE, B., STEFONEK, K., BARNES, B., PATTERSON, J., ZELL, E. R., SCHUCHAT, A. & WHITNEY, C. G. 2001. Epidemiology of invasive Streptococcus pneumoniae infections in the United States, 1995-1998: Opportunities for prevention in the conjugate vaccine era. *Jama,* 285**,** 1729-35.
RODRÍGUEZ, L., CERVANTES, E. & ORTIZ, R. 2011. Malnutrition and Gastrointestinal and Respiratory Infections in Children: A Public Health Problem. *Int J Environ Res Public Health,* 8**,** 1174-205.
RONAGHI, M., KARAMOHAMED, S., PETTERSSON, B., UHLEN, M. & NYREN, P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem,* 242**,** 84-9.
ROSS, R. W. 1939. Acquired tolerance of pneumococcus to M. & B. 693. *Lancet,* 1**,** 1207-1208.
ROSS, W., VRENTAS, C. E., SANCHEZ-VAZQUEZ, P., GAAL, T. & GOURSE, R. L. 2013. The magic spot: a ppGpp binding site on E. coli RNA polymerase responsible for regulation of transcription initiation. *Mol Cell,* 50**,** 420-9.
RUDOLF, D., MICHAYLOV, N., VAN DER LINDEN, M., HOY, L., KLUGMAN, K. P., WELTE, T., PLETZ, M. W. & GROUP, C. S. 2011. International pneumococcal clones match or exceed the fitness of other strains despite the accumulation of antibiotic resistance. *Antimicrob Agents Chemother,* 55**,** 4915-7.
RUIZ, N. 2008. Bioinformatics identification of MurJ (MviN) as the peptidoglycan lipid II flippase in Escherichia coli. *Proc Natl Acad Sci U S A,* 105**,** 15553-7.
RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M. A. & BARRELL, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics,* 16**,** 944-5.
SA-LEAO, R., VILHELMSSON, S. E., DE LENCASTRE, H., KRISTINSSON, K. G. & TOMASZ, A. 2002. Diversity of penicillin-nonsusceptible Streptococcus pneumoniae circulating in Iceland after the introduction of penicillin-resistant clone Spain(6B)-2. *J Infect Dis,* 186**,** 966-75.
SADOWY, E., KUCH, A., GNIADKOWSKI, M. & HRYNIEWICZ, W. 2010a. Expansion and evolution of the Streptococcus pneumoniae Spain9V-ST156 clonal complex in Poland. *Antimicrob Agents Chemother,* 54**,** 1720-7.

SADOWY, E., KUCH, A., GNIADKOWSKI, M. & HRYNIEWICZ, W. 2010b. Expansion and Evolution of the Streptococcus pneumoniae Spain(9V)-ST156 Clonal Complex in Poland. *Antimicrob Agents Chemother,* 54**,** 1720-7.
SAKAI, F., TALEKAR, S. J., KLUGMAN, K. P. & VIDAL, J. E. 2013. Expression of Streptococcus pneumoniae Virulence-Related Genes in the Nasopharynx of Healthy Children. *PLoS One,* 8.
SALTER, S. J., HINDS, J., GOULD, K. A., LAMBERTSEN, L., HANAGE, W. P., ANTONIO, M., TURNER, P., HERMANS, P. W., BOOTSMA, H. J., O'BRIEN, K. L. & BENTLEY, S. D. 2012. Variation at the capsule locus, cps, of mistyped and non-typable Streptococcus pneumoniae isolates. *Microbiology,* 158**,** 1560-9.
SANCAR, A. & REARDON, J. T. 2004. Nucleotide excision repair in E. coli and man. *Adv Protein Chem,* 69**,** 43-71.
SANGER, F. & COULSON, A. R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol,* 94**,** 441-8.
SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-7.
SAUERBIER, J., MAURER, P., RIEGER, M. & HAKENBECK, R. 2012a. Streptococcus pneumoniae R6 interspecies transformation: genetic analysis of penicillin resistance determinants and genome-wide recombination events. *Mol Microbiol,* 86**,** 692-706.
SAUERBIER, J., MAURER, P., RIEGER, M. & HAKENBECK, R. 2012b. Streptococcus pneumoniaeR6 interspecies transformation: genetic analysis of penicillin resistance determinants and genome-wide recombination events. *Molecular Microbiology,* 86**,** 692-706.
SAUVAGE, E., DUEZ, C., HERMAN, R., KERFF, F., PETRELLA, S., ANDERSON, J. W., ADEDIRAN, S. A., PRATT, R. F., FRERE, J. M. & CHARLIER, P. 2007. Crystal structure of the Bacillus subtilis penicillin-binding protein 4a, and its complex with a peptidoglycan mimetic peptide. *J Mol Biol,* 371**,** 528-39.
SAUVAGE, E., KERFF, F., TERRAK, M., AYALA, J. A. & CHARLIER, P. 2008. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiol Rev,* 32**,** 234-58.
SCHADT, E. E., TURNER, S. & KASARSKIS, A. 2010. A window into third-generation sequencing. *Hum Mol Genet,* 19**,** R227-40.
SCHIEMANN, O. & CASPER, W. 1927. Sind die spezifisch pracipitabelen Substanzen der 3 Pneumokokkentypen Haptene. *Zeitschrift fur Hygiene und Infektionskrankheiten,* 108**,** 220-257.
SCHMIDT, L. H., SESLER, C. & DETTWILER, H. A. 1942. Studies on sulfonamide-resistant organisms: I. Development of sulfapyridine resistance by pneumococci. *Journal of Pharmacology and Experimental Therapeutics,* 74**,** 175-189.
SCHMIDT, L. H. & SESLER, C. L. 1943. Development of Resistance to Penicillin by Pneumococci. *Experimental Biology and Medicine,* 52**,** 353-357.
SCHNEERSON, R., ROBBINS, J. B., CHU, C., SUTTON, A., VANN, W., VICKERS, J. C., LONDON, W. T., CURFMAN, B., HARDEGREE, M. C., SHILOACH, J. & ET AL. 1984. Serum antibody responses of juvenile and infant rhesus monkeys injected with Haemophilus influenzae type b and pneumococcus type 6A capsular polysaccharide-protein conjugates. *Infect Immun,* 45**,** 582-91.

SCHNEIDER, G. F. & DEKKER, C. 2012. DNA sequencing with nanopores. *Nat Biotech,* 30**,** 326-328.
SCHUCHAT, A., ROBINSON, K., WENGER, J. D., HARRISON, L. H., FARLEY, M., REINGOLD, A. L., LEFKOWITZ, L. & PERKINS, B. A. 1997. Bacterial meningitis in the United States in 1995. Active Surveillance Team. *N Engl J Med,* 337**,** 970-6.
SCHUSTER, E. F., BLANC, E., PARTRIDGE, L. & THORNTON, J. M. 2007. Correcting for sequence biases in present/absent calls. *Genome Biol,* 8**,** R125.
SCHUSTER, S. C. 2008. Next-generation sequencing transforms today's biology. *Nat Meth,* 5**,** 16-18.
SCHWARTZ, D. C. & CANTOR, C. R. 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell,* 37**,** 67-75.
SCOTT, J. A. 2007. The preventable burden of pneumococcal disease in the developing world. *Vaccine,* 25**,** 2398-405.
SCOTT, J. A., HALL, A. J., DAGAN, R., DIXON, J. M., EYKYN, S. J., FENOLL, A., HORTAL, M., JETTE, L. P., JORGENSEN, J. H., LAMOTHE, F., LATORRE, C., MACFARLANE, J. T., SHLAES, D. M., SMART, L. E. & TAUNAY, A. 1996. Serogroup-specific epidemiology of Streptococcus pneumoniae: associations with age, sex, and geography in 7,000 episodes of invasive disease. *Clin Infect Dis,* 22**,** 973-81.
SELVA, L., VIANA, D., REGEV-YOCHAY, G., TRZCINSKI, K., CORPA, J. M., LASA, I., NOVICK, R. P. & PENADES, J. R. 2009. Killing niche competitors by remote-control bacteriophage induction. *Proc Natl Acad Sci U S A,* 106**,** 1234-8.
SHAK, J. R., VIDAL, J. E. & KLUGMAN, K. P. 2013. Influence of bacterial interactions on pneumococcal colonization of the nasopharynx. *Trends Microbiol,* 21**,** 129-35.
SHENDURE, J., PORRECA, G. J., REPPAS, N. B., LIN, X., MCCUTCHEON, J. P., ROSENBAUM, A. M., WANG, M. D., ZHANG, K., MITRA, R. D. & CHURCH, G. M. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science,* 309**,** 1728-32.
SHI, Z. Y., ENRIGHT, M. C., WILKINSON, P., GRIFFITHS, D. & SPRATT, B. G. 1998. Identification of three major clones of multiply antibiotic-resistant Streptococcus pneumoniae in Taiwanese hospitals by multilocus sequence typing. *J Clin Microbiol,* 36**,** 3514-9.
SHIVSHANKAR, P., SANCHEZ, C., ROSE, L. F. & ORIHUELA, C. J. 2009. The Streptococcus pneumoniae adhesin PsrP binds to Keratin 10 on lung cells. *Mol Microbiol,* 73**,** 663-79.
SIBOLD, C., HENRICHSEN, J., KONIG, A., MARTIN, C., CHALKLEY, L. & HAKENBECK, R. 1994. Mosaic pbpX genes of major clones of penicillin-resistant Streptococcus pneumoniae have evolved from pbpX genes of a penicillin-sensitive Streptococcus oralis. *Mol Microbiol,* 12**,** 1013-23.
SIERRA, A., LOPEZ, P., ZAPATA, M., VANEGAS, B., CASTREJON, M., DEANTONIO, R., HAUSDORFF, W. & COLINDRES, R. 2011. Non-typeable Haemophilus influenzae and Streptococcus pneumoniae as primary causes of acute otitis media in colombian children: a prospective study. *BMC Infectious Diseases,* 11**,** 4.
SILVA. 2015. *silva: high quality ribosomal RNA databases* [Online]. Available: http://www.arb-silva.de/documentation/faqs/ [Accessed 23 July 2015].

SIMBERKOFF, M. S., EL SADR, W., SCHIFFMAN, G. & RAHAL, J. J., JR. 1984. Streptococcus pneumoniae infections and bacteremia in patients with acquired immune deficiency syndrome, with report of a pneumococcal vaccine failure. *Am Rev Respir Dis,* 130**,** 1174-6.
SIMELL, B., AURANEN, K., KAYHTY, H., GOLDBLATT, D., DAGAN, R. & O'BRIEN, K. L. 2012. The fundamental link between pneumococcal carriage and disease. *Expert Rev Vaccines,* 11**,** 841-55.
SIMONSEN, L., FUKUDA, K., SCHONBERGER, L. B. & COX, N. J. 2000. The impact of influenza epidemics on hospitalizations. *J Infect Dis,* 181**,** 831-7.
SINGER, C. 1928. A Short History of Medicine. *Ancient Greece.* New York: Oxfird University Press.
SINGH, V. K., MORE, T. & SINGH, S. 1997. The effect of activation of granulocytes on enzyme release and hydrogen peroxide and superoxide production in buffaloes. *Vet Res Commun,* 21**,** 241-7.
SLATKIN, M. 2008. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet,* 9**,** 477-85.
SLEEMAN, K. L., GRIFFITHS, D., SHACKLEY, F., DIGGLE, L., GUPTA, S., MAIDEN, M. C., MOXON, E. R., CROOK, D. W. & PETO, T. E. 2006. Capsular serotype-specific attack rates and duration of carriage of Streptococcus pneumoniae in a population of children. *J Infect Dis,* 194**,** 682-8.
SMITH, A. M. & KLUGMAN, K. P. 1997. Three predominant clones identified within penicillin-resistant South African isolates of Streptococcus pneumoniae. *Microb Drug Resist,* 3**,** 385-9.
SMITH, A. M. & KLUGMAN, K. P. 1998. Alterations in PBP 1A essential-for high-level penicillin resistance in Streptococcus pneumoniae. *Antimicrob Agents Chemother,* 42**,** 1329-33.
SMITH, A. M. & KLUGMAN, K. P. 2001. Alterations in MurM, a Cell Wall Muropeptide Branching Enzyme, Increase High-Level Penicillin and Cephalosporin Resistance in Streptococcus pneumoniae. *Antimicrob Agents Chemother,* 45**,** 2393-6.
SMITH, G. R. 2012. How RecBCD enzyme and Chi promote DNA break repair and recombination: a molecular biologist's view. *Microbiol Mol Biol Rev,* 76**,** 217-28.
SONG, X.-M., CONNOR, W., HOKAMP, K., BABIUK, L. A. & POTTER, A. A. 2009. Transcriptome studies on Streptococcus pneumoniae, illustration of early response genes to THP-1 human macrophages. *Genomics,* 93**,** 72-82.
SPRATT, B. G. & GREENWOOD, B. M. 2000. Prevention of pneumococcal disease by vaccination: does serotype replacement matter? *Lancet,* 356**,** 1210-1.
STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30**,** 1312-3.
STAMATAKIS, A., LUDWIG, T. & MEIER, H. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics,* 21**,** 456-63.
STEENHOFF, A. P., SHAH, S. S., RATNER, A. J., PATIL, S. M. & MCGOWAN, K. L. 2006. Emergence of Vaccine-Related Pneumococcal Serotypes as a Cause of Bacteremia. *Clinical Infectious Diseases,* 42**,** 907-914.
STEINHOFF, M. C., EDWARDS, K., KEYSERLING, H., THOMS, M. L., JOHNSON, C., MADORE, D. & HOGERMAN, D. 1994. A randomized comparison of three bivalent Streptococcus pneumoniae glycoprotein conjugate vaccines in

young children: effect of polysaccharide size and linkage characteristics. *Pediatr Infect Dis J,* 13**,** 368-72.
STERNBERG, G. M. 1881. A fatal form of septicaemia in the rabbit, produced by the subcutaneous injection of human saliva. *National Board of Health Bulletin,* 2.
STERNBERG, G. M. 1982. A fatal form of septicaemia in the rabbit, produced by the subcutaneous injection of human saliva. *Studies from the Biological Laboratory of Johns Hopkins University,* 2.
STEVENS, K. E., CHANG, D., ZWACK, E. E. & SEBERT, M. E. 2011. Competence in Streptococcus pneumoniae is regulated by the rate of ribosomal decoding errors. *MBio,* 2.
STOHS, S. J. & BAGCHI, D. 1995. Oxidative mechanisms in the toxicity of metal ions. *Free Radic Biol Med,* 18**,** 321-36.
STOLTZFUS, A., LESLIE, J. F. & MILKMAN, R. 1988. Molecular evolution of the Escherichia coli chromosome. I. Analysis of structure and natural variation in a previously uncharacterized region between trp and tonB. *Genetics,* 120**,** 345-58.
SWIATLO, E., CHAMPLIN, F. R., HOLMAN, S. C., WILSON, W. W. & WATT, J. M. 2002. Contribution of Choline-Binding Proteins to Cell Surface Properties of Streptococcus pneumoniae. *Infect Immun,* 70**,** 412-5.
TAHA, T. E., DALLABETTA, G. A., HOOVER, D. R., CHIPHANGWI, J. D., MTIMAVALYE, L. A., LIOMBA, G. N., KUMWENDA, N. I. & MIOTTI, P. G. 1998. Trends of HIV-1 and sexually transmitted diseases among pregnant and postpartum women in urban Malawi. *Aids,* 12**,** 197-203.
TAKADA, K., HAYASHI, K., SASAKI, K., SATO, T. & HIRASAWA, M. 2006. Selectivity of Mitis Salivarius agar and a new selective medium for oral streptococci in dogs. *J Microbiol Methods,* 66**,** 460-5.
TAKAO, A., NAGAMUNE, H. & MAEDA, N. 2004. Identification of the anginosus group within the genus Streptococcus using polymerase chain reaction. *FEMS Microbiol Lett,* 233**,** 83-9.
TANIAI, H., IIDA, K., SEKI, M., SAITO, M., SHIOTA, S., NAKAYAMA, H. & YOSHIDA, S. 2008. Concerted action of lactate oxidase and pyruvate oxidase in aerobic growth of Streptococcus pneumoniae: role of lactate as an energy source. *J Bacteriol,* 190**,** 3572-9.
TANO, K., GRAHN-HAKANSSON, E., HOLM, S. E. & HELLSTROM, S. 2000. Inhibition of OM pathogens by alpha-hemolytic streptococci from healthy children, children with SOM and children with rAOM. *Int J Pediatr Otorhinolaryngol,* 56**,** 185-90.
TAPLIN, G. V., JACOX, R. F. & HOWLAND, J. W. 1940. THe use of sodium sulfapyridine by hypodermoclysis. *Journal of the American Medical Association,* 114**,** 1733-1734.
TENG, L. J., HSUEH, P. R., CHEN, Y. C., HO, S. W. & LUH, K. T. 1998. Antimicrobial susceptibility of viridans group streptococci in Taiwan with an emphasis on the high rates of resistance to penicillin and macrolides in Streptococcus oralis. *J Antimicrob Chemother,* 41**,** 621-7.
TETTELIN, H., MASIGNANI, V., CIESLEWICZ, M. J., DONATI, C., MEDINI, D., WARD, N. L., ANGIUOLI, S. V., CRABTREE, J., JONES, A. L., DURKIN, A. S., DEBOY, R. T., DAVIDSEN, T. M., MORA, M., SCARSELLI, M., MARGARIT Y ROS, I., PETERSON, J. D., HAUSER, C. R., SUNDARAM, J. P., NELSON, W. C.,

MADUPU, R., BRINKAC, L. M., DODSON, R. J., ROSOVITZ, M. J., SULLIVAN, S. A., DAUGHERTY, S. C., HAFT, D. H., SELENGUT, J., GWINN, M. L., ZHOU, L., ZAFAR, N., KHOURI, H., RADUNE, D., DIMITROV, G., WATKINS, K., O'CONNOR, K. J., SMITH, S., UTTERBACK, T. R., WHITE, O., RUBENS, C. E., GRANDI, G., MADOFF, L. C., KASPER, D. L., TELFORD, J. L., WESSELS, M. R., RAPPUOLI, R. & FRASER, C. M. 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A,* 102**,** 13950-5.

TETTELIN, H., NELSON, K. E., PAULSEN, I. T., EISEN, J. A., READ, T. D., PETERSON, S., HEIDELBERG, J., DEBOY, R. T., HAFT, D. H., DODSON, R. J., DURKIN, A. S., GWINN, M., KOLONAY, J. F., NELSON, W. C., PETERSON, J. D., UMAYAM, L. A., WHITE, O., SALZBERG, S. L., LEWIS, M. R., RADUNE, D., HOLTZAPPLE, E., KHOURI, H., WOLF, A. M., UTTERBACK, T. R., HANSEN, C. L., MCDONALD, L. A., FELDBLYUM, T. V., ANGIUOLI, S., DICKINSON, T., HICKEY, E. K., HOLT, I. E., LOFTUS, B. J., YANG, F., SMITH, H. O., VENTER, J. C., DOUGHERTY, B. A., MORRISON, D. A., HOLLINGSHEAD, S. K. & FRASER, C. M. 2001. Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. *Science,* 293**,** 498-506.

THERMO SCIENTIFIC. 2015. *Brain Heart Infusion Broth* [Online]. Available: http://www.oxoid.com/UK/blue/prod_detail/prod_detail.asp?pr=CM1135&org=133 [Accessed 7 July 2015].

THORNTON, G. F. & ANDRIOLE, V. T. 1966. Laboratory and clinical studies of a new antibiotic, cephaloridine, in the treatment of gram-positive infections. *Yale J Biol Med,* 39**,** 9-20.

THROUP, J. P., KORETKE, K. K., BRYANT, A. P., INGRAHAM, K. A., CHALKER, A. F., GE, Y., MARRA, A., WALLIS, N. G., BROWN, J. R., HOLMES, D. J., ROSENBERG, M. & BURNHAM, M. K. 2000. A genomic analysis of two-component signal transduction in Streptococcus pneumoniae. *Mol Microbiol,* 35**,** 566-76.

TILLETT, W. S., CAMBIER, M. J. & HARRIS, W. H. 1943. Sulfonamide-fast pneumoccoi. A clinical report of two cases of pneumonia together with experimental studies on the effectiveness of penicillin and tyrothricin against sulfonamide-resistant strains. *J Clin Invest,* 22**,** 249-55.

TILLETT, W. S., CAMBIER, M. J. & MCCORMACK, J. E. 1944. The Treatment of Lobar Pneumonia and Pneumococcal Empyema with Penicillin. *Bull N Y Acad Med,* 20**,** 142-78.

TIPPER, D. J. & STROMINGER, J. L. 1965. Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. *Proc Natl Acad Sci U S A,* 54**,** 1133-41.

TOMASZ, A. 1979. The mechanism of the irreversible antimicrobial effects of penicillins: how the beta-lactam antibiotics kill and lyse bacteria. *Annu Rev Microbiol,* 33**,** 113-37.

TOMASZ, A. 1997. Antibiotic resistance in Streptococcus pneumoniae. *Clin Infect Dis,* 24 Suppl 1**,** S85-8.

TOMASZ, A., ALBINO, A. & ZANATI, E. 1970. Multiple antibiotic resistance in a bacterium with suppressed autolytic system. *Nature,* 227**,** 138-40.

TOMASZ, A., MOREILLON, P. & POZZI, G. 1988. Insertional inactivation of the major autolysin gene of Streptococcus pneumoniae. *J Bacteriol,* 170**,** 5931-4.

TONG, H., CHEN, W., MERRITT, J., QI, F., SHI, W. & DONG, X. 2007. Streptococcus oligofermentans inhibits Streptococcus mutans through conversion of lactic acid into inhibitory H2O2: a possible counteroffensive strategy for interspecies competition. *Mol Microbiol,* 63**,** 872-80.
TOVPEKO, Y. & MORRISON, D. A. 2014. Competence for genetic transformation in Streptococcus pneumoniae: mutations in sigmaA bypass the comW requirement. *J Bacteriol,* 196**,** 3724-34.
TROTMAN, J., HUGHES, B. & MOLLISON, L. 1995. Invasive pneumococcal disease in central Australia. *Clin Infect Dis,* 20**,** 1553-6.
TROTTER, C. L. & GREENWOOD, B. M. 2007. Meningococcal carriage in the African meningitis belt. *Lancet Infect Dis,* 7**,** 797-803.
TRZCINSKI, K., THOMPSON, C. M., GILBEY, A. M., DOWSON, C. G. & LIPSITCH, M. 2006. Incremental increase in fitness cost with increased beta -lactam resistance in pneumococci evaluated by competition in an infant rat nasal colonization model. *J Infect Dis,* 193**,** 1296-303.
TSUI, H.-C. T., BOERSMA, M. J., VELLA, S. A., KOCAOGLU, O., KURU, E., PECENY, J. K., CARLSON, E. E., VANNIEUWENHZE, M. S., BRUN, Y. V., SHAW, S. L. & WINKLER, M. E. 2014. Pbp2x localizes separately from Pbp2b and other peptidoglycan synthesis proteins during later stages of cell division of Streptococcus pneumoniae D39. *Molecular Microbiology,* 94**,** 21-40.
TU, A. H., FULGHAM, R. L., MCCRORY, M. A., BRILES, D. E. & SZALAI, A. J. 1999. Pneumococcal surface protein A inhibits complement activation by Streptococcus pneumoniae. *Infect Immun,* 67**,** 4720-4.
TUOMANEN, E. I. 1997. The Biology of Pneumococcal Infection. *Pediatr Res,* 42**,** 253-258.
TURNER, P., TURNER, C., JANKHOT, A., HELEN, N., LEE, S. J., DAY, N. P., WHITE, N. J., NOSTEN, F. & GOLDBLATT, D. 2012. A longitudinal study of Streptococcus pneumoniae carriage in a cohort of infants and their mothers on the Thailand-Myanmar border. *PLoS One,* 7**,** e38271.
USUF, E., BOTTOMLEY, C., ADEGBOLA, R. A. & HALL, A. 2014. Pneumococcal carriage in sub-Saharan Africa--a systematic review. *PLoS One,* 9**,** e85001.
VAN DE PEER, Y., CHAPELLE, S. & DE WACHTER, R. 1996. A quantitative map of nucleotide substitution rates in bacterial rRNA. *Nucleic Acids Res,* 24**,** 3381-91.
VAN ROSSUM, A. M., LYSENKO, E. S. & WEISER, J. N. 2005. Host and bacterial factors contributing to the clearance of colonization by Streptococcus pneumoniae in a murine model. *Infect Immun,* 73**,** 7718-26.
VEENHOVEN, R., BOGAERT, D., UITERWAAL, C., BROUWER, C., KIEZEBRINK, H., BRUIN, J., E, I. J., HERMANS, P., DE GROOT, R., ZEGERS, B., KUIS, W., RIJKERS, G., SCHILDER, A. & SANDERS, E. 2003. Effect of conjugate pneumococcal vaccine followed by polysaccharide pneumococcal vaccine on recurrent acute otitis media: a randomised study. *Lancet,* 361**,** 2189-95.
VERSCHOOR, C. P., JOHNSTONE, J., LOEB, M., BRAMSON, J. L. & BOWDISH, D. M. 2014. Anti-pneumococcal deficits of monocyte-derived macrophages from the advanced-age, frail elderly and related impairments in PI3K-AKT signaling. *Hum Immunol,* 75**,** 1192-6.
VINELLA, D., JOSELEAU-PETIT, D., THEVENET, D., BOULOC, P. & D'ARI, R. 1993. Penicillin-binding protein 2 inactivation in Escherichia coli results in cell

division inhibition, which is relieved by FtsZ overexpression. *J Bacteriol,* 175**,** 6704-10.
VINJE, H., ALMOY, T., LILAND, K. H. & SNIPEN, L. 2014. A systematic search for discriminating sites in the 16S ribosomal RNA gene. *Microb Inform Exp,* 4**,** 2.
VOS, M. 2009. Why do bacteria engage in homologous recombination? *Trends Microbiol,* 17**,** 226-32.
WAHL, B., ERNST, F., KUMAR, Y., MÜLLER, B., STANGIER, K. & PAPROTKA, T. Defining the Microbial Composition of Environmental Samples Using Next Generation Sequencing.
WAITE, R. D. 2001. *The molecular genetic basis of biofilm generated capsule phase variation in Streptococcus pneumoniae.* Doctor of Philosophy, University of Warwick.
WAITE, R. D., PENFOLD, D. W., STRUTHERS, J. K. & DOWSON, C. G. 2003. Spontaneous sequence duplications within capsule genes cap8E and tts control phase variation in Streptococcus pneumoniae serotypes 8 and 37. *Microbiology,* 149**,** 497-504.
WAITE, R. D., STRUTHERS, J. K. & DOWSON, C. G. 2001. Spontaneous sequence duplication within an open reading frame of the pneumococcal type 3 capsule locus causes high-frequency phase variation. *Mol Microbiol,* 42**,** 1223-32.
WALKER, C. L., RUDAN, I., LIU, L., NAIR, H., THEODORATOU, E., BHUTTA, Z. A., O'BRIEN, K. L., CAMPBELL, H. & BLACK, R. E. 2013. Global burden of childhood pneumonia and diarrhoea. *Lancet,* 381**,** 1405-16.
WALL, E. C., EVERETT, D. B., MUKAKA, M., BAR-ZEEV, N., FEASEY, N., JAHN, A., MOORE, M., VAN OOSTERHOUT, J. J., PENSALO, P., BAGUIMIRA, K., GORDON, S. B., MOLYNEUX, E. M., CARROL, E. D., FRENCH, N., MOLYNEUX, M. E. & HEYDERMAN, R. S. 2014. Bacterial meningitis in Malawian adults, adolescents, and children during the era of antiretroviral scale-up and Haemophilus influenzae type b vaccination, 2000-2012. *Clin Infect Dis,* 58**,** e137-45.
WANI, J. H., GILBERT, J. V., PLAUT, A. G. & WEISER, J. N. 1996. Identification, cloning, and sequencing of the immunoglobulin A1 protease gene of Streptococcus pneumoniae. *Infect Immun,* 64**,** 3967-74.
WARD, J. I., BERKOWITZ, C., BURKART, K., BRENNEMAN, G., PESCETTI, J. & MARCY, M. 1988. Comparative immunogenicity of H. influenzae type b (Hib) conjugate vaccines (PRP-D/PRP-OMP) in infants less than 6 months of age. *Twenty-Eighth Interscience Conference of Antimicrobial Agents and Chemotherapy.* Los Angeles, California.
WASHBOURN, J. W. 1897. Antipneumococic serum. *British Medical Journal,* 1887.
WATSON, D. A., MUSHER, D. M., JACOBSON, J. W. & VERHOEF, J. 1993. A brief history of the pneumococcus in biomedical research: a panoply of scientific discovery. *Clin Infect Dis,* 17**,** 913-24.
WATSON, K., CARVILLE, K., BOWMAN, J., JACOBY, P., RILEY, T. V., LEACH, A. J. & LEHMANN, D. 2006. Upper respiratory tract bacterial carriage in Aboriginal and non-Aboriginal children in a semi-arid area of Western Australia. *Pediatr Infect Dis J,* 25**,** 782-90.

WEEN, O., GAUSTAD, P. & HAVARSTEIN, L. S. 1999. Identification of DNA binding sites for ComE, a key regulator of natural competence in Streptococcus pneumoniae. *Mol Microbiol,* 33**,** 817-27.
WEICHSELBAUM, A. 1886a. Aetiologie und pathologische Anatomie der akuten Lungenentziindungen. *Wiener Medizinische Wochenschrift,* 36.
WEICHSELBAUM, A. 1886b. Aetiologie und pathologische Anatomie der akuten Lungenentziindungen. *Wiener Medizinische Wochenschrift,* 36.
WEICHSELBAUM, A. 1986. Ueber die Aetiologie der acuten Lungenund Rippenfellentzundungen. *Med. Jahrb., 1, NF.*
WEINBERGER, D. M., MALLEY, R. & LIPSITCH, M. 2011. Serotype replacement in disease following pneumococcal vaccination: A discussion of the evidence. *Lancet,* 378**,** 1962-73.
WEINSTEIN, M. P., KLUGMAN, K. P. & JONES, R. N. 2009. Rationale for revised penicillin susceptibility breakpoints versus Streptococcus pneumoniae: coping with antimicrobial susceptibility in an era of resistance. *Clin Infect Dis,* 48**,** 1596-600.
WEISER, J. N., BAE, D., EPINO, H., GORDON, S. B., KAPOOR, M., ZENEWICZ, L. A. & SHCHEPETOV, M. 2001. Changes in Availability of Oxygen Accentuate Differences in Capsular Polysaccharide Expression by Phenotypic Variants and Clinical Isolates of Streptococcus pneumoniae. *Infect Immun,* 69**,** 5430-9.
WELTE, T. 2012. Risk factors and severity scores in hospitalized patients with community-acquired pneumonia: prediction of severity and mortality. *Eur J Clin Microbiol Infect Dis,* 31**,** 33-47.
WESCOMBE, P. A., UPTON, M., RENAULT, P., WIRAWAN, R. E., POWER, D., BURTON, J. P., CHILCOTT, C. N. & TAGG, J. R. 2011. Salivaricin 9, a new lantibiotic produced by Streptococcus salivarius. *Microbiology,* 157**,** 1290-9.
WETZEL, J., KINGSFORD, C. & POP, M. 2011. Assessing the benefits of using mate-pairs to resolve repeats in de novo short-read prokaryotic assemblies. *BMC Bioinformatics,* 12**,** 95.
WHATMORE, A. M., BARCUS, V. A. & DOWSON, C. G. 1999. Genetic diversity of the streptococcal competence (com) gene locus. *J Bacteriol,* 181**,** 3144-54.
WHATMORE, A. M., EFSTRATIOU, A., PICKERILL, A. P., BROUGHTON, K., WOODARD, G., STURGEON, D., GEORGE, R. & DOWSON, C. G. 2000. Genetic relationships between clinical isolates of Streptococcus pneumoniae, Streptococcus oralis, and Streptococcus mitis: characterization of "Atypical" pneumococci and organisms allied to S. mitis harboring S. pneumoniae virulence factor-encoding genes. *Infect Immun,* 68**,** 1374-82.
WHILEY, R. A., FRASER, H., HARDIE, J. M. & BEIGHTON, D. 1990. Phenotypic differentiation of Streptococcus intermedius, Streptococcus constellatus, and Streptococcus anginosus strains within the "Streptococcus milleri group". *J Clin Microbiol,* 28**,** 1497-501.
WHITBY, L. E. H. 1938. Chemotherapy of pneumococcal and other infections - With 2-(p-aminobenzenesulphonamido) pyridine. *Lancet,* 1**,** 1210-1212.
WHITE, A. N., NG, V., SPAIN, C. V., JOHNSON, C. C., KINLIN, L. M. & FISMAN, D. N. 2009. Let the sun shine in: effects of ultraviolet radiation on invasive pneumococcal disease risk in Philadelphia, Pennsylvania. *BMC Infect Dis,* 9**,** 196.

WHITE, B. 1938. *The Biology of Pneumococcus.,* New York, Oxford University Press.
WHITNEY, C. G., FARLEY, M. M., HADLER, J., HARRISON, L. H., BENNETT, N. M., LYNFIELD, R., REINGOLD, A., CIESLAK, P. R., PILISHVILI, T., JACKSON, D., FACKLAM, R. R., JORGENSEN, J. H. & SCHUCHAT, A. 2003. Decline in Invasive Pneumococcal Disease after the Introduction of Protein–Polysaccharide Conjugate Vaccine. *New England Journal of Medicine,* 348**,** 1737-1746.
WHO 2000. Detecting meningococcal meningitis epidemics in highly-endemic African countries. *Wkly Epidemiol Rec,* 75**,** 306-9.
WHO. 2015. *Pneumococcal disease* [Online]. Available: http://www.who.int/ith/diseases/pneumococcal/en/ [Accessed 6 July 2015].
WIEGAND, I., HILPERT, K. & HANCOCK, R. E. 2008. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat Protoc,* 3**,** 163-75.
WILLCOX, M. D., ZHU, H. & KNOX, K. W. 2001. Streptococcus australis sp. nov., a novel oral streptococcus. *Int J Syst Evol Microbiol,* 51**,** 1277-81.
WILLIAMS, R., PEISAJOVICH, S. G., MILLER, O. J., MAGDASSI, S., TAWFIK, D. S. & GRIFFITHS, A. D. 2006. Amplification of complex gene libraries by emulsion PCR. *Nat Methods,* 3**,** 545-50.
WOESE, C. R. 1987. Bacterial evolution. *Microbiol Rev,* 51**,** 221-71.
WRIGHT, A., PARRY MORGAN, W., COLEBROOK, L. & DODGSON, R. W. Observations on prophylactic inoculation against pneumococcus infections, and on the results which have been achieved by it. *The Lancet,* 183**,** 1-10.
YANG, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci,* 13**,** 555-6.
YE, K., SCHULZ, M. H., LONG, Q., APWEILER, R. & NING, Z. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics,* 25**,** 2865-71.
YEATS, C., FINN, R. D. & BATEMAN, A. 2002. The PASTA domain: a beta-lactam-binding domain. *Trends Biochem Sci,* 27**,** 438.
YESILKAYA, H., ANDISI, V. F., ANDREW, P. W. & BIJLSMA, J. J. 2013. Streptococcus pneumoniae and reactive oxygen species: an unusual approach to living with radicals. *Trends Microbiol,* 21**,** 187-95.
YESILKAYA, H., KADIOGLU, A., GINGLES, N., ALEXANDER, J. E., MITCHELL, T. J. & ANDREW, P. W. 2000. Role of Manganese-Containing Superoxide Dismutase in Oxidative Stress and Virulence of Streptococcus pneumoniae. *Infect Immun,* 68**,** 2819-26.
ZAPUN, A., CONTRERAS-MARTEL, C. & VERNET, T. 2008. Penicillin-binding proteins and beta-lactam resistance. *FEMS Microbiol Rev,* 32**,** 361-85.
ZERBINO, D. R. & BIRNEY, E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res,* 18**,** 821-9.
ZERFASS, I., HAKENBECK, R. & DENAPAITE, D. 2009. An important site in PBP2x of penicillin-resistant clinical isolates of Streptococcus pneumoniae: mutational analysis of Thr338. *Antimicrob Agents Chemother,* 53**,** 1107-15.

ZHANG, L., LI, Z., WAN, Z., KILBY, A., KILBY, J. M. & JIANG, W. 2015. Humoral Immune Responses to Streptococcus pneumoniae in the Setting of HIV-1 Infection. *Vaccine,* 33**,** 4430-4436.
ZHANG, Z., CLARKE, T. B. & WEISER, J. N. 2009. Cellular effectors mediating Th17-dependent clearance of pneumococcal colonization in mice. *J Clin Invest,* 119**,** 1899-909.
ZHOU, X. & STEPHENS, M. 2014. Efficient Algorithms for Multivariate Linear Mixed Models in Genome-wide Association Studies. *Nat Methods,* 11**,** 407-9.
ZIGHELBOIM, S. & TOMASZ, A. 1980. Penicillin-binding proteins of multiply antibiotic-resistant South African strains of Streptococcus pneumoniae. *Antimicrobial Agents and Chemotherapy,* 17**,** 434-442.
ZUMBO, P. 2012. *Ethanol Precipitation* [Online]. Available: http://physiology.med.cornell.edu/faculty/mason/lab/zumbo/files/ETHANOL_PRECIPITATION.pdf [Accessed 8 July 2015].

10 APPENDIX

10.1 Blood culture SOP

Malawi-Liverpool-Wellcome

Clinical Research Programme

STANDARD OPERATING PROCEDURE	MLW.SOP.C.8.000 Version 1 *X = insert departmental code* *Y = insert section code* *Z= insert version number*
Obtaining and clinical processing of blood cultures in adult medicine	
Effective date: 1/6/2012	Next Review Due: 31/5/2013

	Name and Position	Signature	Date
Author	Dr Emma Wall, PI BAM study Vella Kadzu Clinical nurse manager MLW		23/5/2012
Reviewer	Mike Moore, laboratory manager Brigitte Denis deputy laboratory manager Dr Mulinda Nyrienda, Head of department, AETC Dr Jane Mallewa, consultant physician, College of Medicine		23/5/2012
Approver	Professor Rob Heyderman, Director	le	

Revision History

Ver. No.	Revised by (Name & Position)	Effective Date	Details of changes

1. **Background**

A blood culture is a test to detect blood stream infection. One may be ordered when a patient has any of the following systematic symptoms of an infection: high fever, rigors, night sweats or chills and the clinician suspects bacteria or fungi may be present in the blood. The culture can disclose what type of pathogen is causing the infection, which will determine how it is treated.

2. **Purpose**

The purpose of this SOP is to describe procedures once a blood culture has been collected in adult medicine. Adherence to this SOP will ensure uniformity of specimen management from the patient to the laboratory and to ensure the results are available to the clinical team and the patient where indicated as soon as they are available. The adult blood culture service at QECH is currently only for the use of medical patients.

3. **Responsibility**

3.1 The guideline is intended for all MLW and QECH medical and nursing staff taking adult blood cultures for microbiology in QECH to be processed by clinical microbiology at MLW.

3.2 The member of staff collecting the blood culture is responsible for ensuring that the appropriate MLW specimen request form and blood culture bottle are correctly and accurately completed and that they are sent to the MLW lab as soon as possible after the specimen has been collected.

3.3 Only the on call medical team can identify patients requiring blood culture in the AETC.

3.4 On the medical wards it is the responsibility of the clinician covering the relevant bay to identify patients requiring blood culture at any given time.

3.5 It is the responsibility of the ward sisters to ensure that only competent and trained staffs collect blood culture specimens and that untrained staff do not collect blood cultures until trained and have been deemed competent to do so.

3.6 During working hours it is the responsibility of the blood culture courier to ensure all samples are received by the laboratory in a timely manner.

3.7 MLW provides materials, equipment, and technical training to staff involved in collecting blood cultures at QECH adult medical wards.

3.8 It is the responsibility of the blood culture nurses to ensure that adequate stock of equipment for blood cultures is left in the AETC for each weekend. When stocks are running low the blood culture nurses must replenish the stocks and remove expired stock including collection bottles. The blood culture nurses or courier can requisition further supplies from MLW stores/laboratory reception.

3.9 It is the responsibility of the AETC team leader for each shift to note the number of blood culture bottles available in the AETC and to ensure these bottles are kept securely for use within the AETC and **not** given to other wards. The team leader must also ensure that at weekends blood culture specimens are batched in the collection box at the staff base, and taken to the laboratory reception at the paediatric research ward at the following times: 7am, 11am, 2pm and 4pm.

3.10 Each nurse in charge on the medical wards must keep a note of the number of blood culture bottles available and ensure adequate stocks are kept at all times. Bottles must not be taken from the AETC when ward stocks are low.

3.11 The MLW laboratory must ensure that all culture results are printed and delivered to the relevant ward in a timely manner.

3.12 It is the responsibility of the doctors and nurses in charge of the patient to ensure that blood culture results are filed in the patient's notes, and any appropriate treatment based on the results is prescribed.

4. Training

4.1 All staff obtaining blood culture specimens must receive training arranged by MLW through the QECH ward in-charges and/or departmental heads. The course is delivered by the MLW Clinical Manager. Both nurses and medical staff should be trained.

4.2 Regular training sessions will be held quarterly.CPD points can be awarded for attending training for nurses and clinical officers. All clinical officers and interns should be trained in obtaining blood cultures.

4.3 Refresher sessions will be held annually for trained staff and it is the responsibility of trained staff to ensure they are kept up to date.

5. Procedure

5.1 Identifying patients for blood culture.

Blood cultures should be taken from patients who meet the following criteria:

- a fever of >37.5 or <35 degrees centigrade where a clinical suspicion of bacteraemia exists, including hospitalised patients with leucopoenia or who are immunocompromised.
- Blood cultures should always be ordered in patients where a clinical suspicion of sepsis, meningitis, pneumonia, endocarditis, osteomyelitis, abdominal abscess, severe skin and soft tissue infections, dialysis catheter infection, fever of unknown origin and septic arthritis exists. In cases of severe malaria in adults a blood culture may be considered after discussion with a senior clinician.
- Blood cultures should only be taken in patients who are being admitted to hospital, or are under regular close review by a consultant physician (e.g. dialysis and haematology patients).
- Blood culture should ideally be taken pre-antibiotic treatment.

5.2 Locations for blood culture.

5.2.1 Patients in the AETC ward should have blood culture drawn only if they meet the clinical criteria in 5.1 and are going to be admitted under adult medicine. Medical inpatients should have blood culture drawn if they meet the clinical criteria in 5.1

5.2.2 Medical in-patients who have received intravenous antibiotics should be discussed with the registrar or consultant in charge of that patient before blood cultures are taken.

5.3 Collection of Patient Specimen

5.3.1 Please refer to SOP 'C001 Obtaining blood culture' available on the MLW intranet.

5.4 Sample processing

5.4.1 Label the blood culture bottle with the full name of the patient, the ward he/she is admitted to and the time and date the blood was collected. If the specimen is taken in the AETC and the destination ward is known, **please document the destination ward only on both the bottle and the request form**. This will ensure the results are sent to the correct ward. Put the bottle in the specimen bag.

5.4.2 Complete the MLW Lab request form and place the form in the designated pouch in the specimen bag.

5.4.3 Record in the patient's file, the date and time the blood culture specimen was collected, the volume collected and the arm (site) venepuncture was performed. Put your signature and job title.

5.4.4 Send the specimen and the form to the MLW Specimen Reception at the Paediatric Research Ward Laboratory as soon as possible via the MLW blood culture courier in working hours.

Out of hours

5.4.5 All blood culture specimens taken at night should be placed in the blood culture sample collection box located at the staff base in AETC for collection by the courier the following morning. The MLW laboratory has deemed this a safe procedure that will not result in early death or damage to micro-organisms.

5.4.6 During weekends blood culture specimens taken in the AETC should be batched in the sample collection box in AETC. The AETC shift leader will ensure that a designated member of staff takes the blood cultures to the laboratory four times per day, at 7am, 11am, 2pm and 4pm. If no specimen receptionist is available, the specimen should be left in the collection box.

5.4.7 Blood cultures taken on the wards at weekends should be delivered before 4pm to the MLW specimen reception by the member of staff taking the specimen.

5.5 Obtaining results

5.5.1 All positive blood culture results are released daily by the MLW laboratory. Monday to Friday these are taken by the MLW laboratory messenger to the ward indicated on the request form. Positive results released at the weekend will be placed in the appropriate ward pigeon hole outside the laboratory reception at the paediatric research ward (PRW). Where no ward is indicated, the results will be placed in the PRW pigeon hole labelled 'public'.

5.5.2 All results are printed and released by the laboratory after 5 days of culture. These results are sent to the relevant ward listed on the request form. Negative results are also printed and sent to the relevant ward. Results are also available online via the Laboratory Information Management System (LIMS).

5.5.3 Results where only AETC was listed as the location without a destination ward will be delivered by the laboratory to the most likely destination ward for that patient (e.g. 3B for adult male patients and 4A for adult female patients. **Listing AETC only as the location for the blood culture is therefore likely to lead to a delay in results being picked up by the clinicians caring for that patient and should be avoided.**

5.5.4 **Results will not be given directly to patients by the laboratory staff. Patients who are discharged and advised to return to collect their results sho**uld be advised to attend the ward of discharge to obtain their results from the treating clinical team.

6.0 Health and Safety

Blood is a body fluid that may be infected with viruses circulating during acute and chronic viral infection. These viruses include HIV, hepatitis B and C. If you are exposed to blood via a needlestick injury or eye splash while either taking blood for culture or handling blood specimens, you may be at risk of contracting viral infection. You must immediately follow the procedures in your department for management of a needlestick injury and inform your line manager.

7.0 Associated Procedures

8.0 **Reference**

10.2 Lumbar puncture SOP

Malawi-Liverpool-Wellcome

Clinical Research Programme

STANDARD OPERATING PROCEDURE	MLW.SOP.X.Y.000 Version Z *X = insert departmental code* *Y = insert section code* *Z= insert version number*
Lumbar puncture	
Effective date: 2/3/2012	Next Review Due: 2/3/2013

	Name and Position	Signature	Date
Author	Dr Emma Wall, PI		24/2/2012
Reviewer			
Approver			

Revision History

Ver. No.	Revised by (Name & Position)	Effective Date	Details of changes

6. Background

A lumbar puncture is a procedure to obtain cerebrospinal fluid (CSF) culture to detect infection in the fluid surrounding the meningeal layers covering the brain and spinal cord. During a lumbar puncture a needle is inserted into the fluid space around the spinal cord and fluid is removed. One may be ordered when a patient has symptoms of a meningeal infection; such as a high fever, severe headache, neck stiffness, photophobia (inability to tolerate bright light), confusion, convulsions or coma, and the clinician suspects bacteria, fungi or viruses have spread into the CSF or brain. The culture can disclose what type of pathogen is causing the infection, which will determine how it is treated.

7. Purpose

7.1 To ensure standardized lumbar puncture procedures for patient safety and to minimize sample contamination.

8. Responsibility

8.1 It is the responsibility of the supervising clinician to ensure all lumbar punctures (LP's) are performed safely and appropriately on patients in their care.

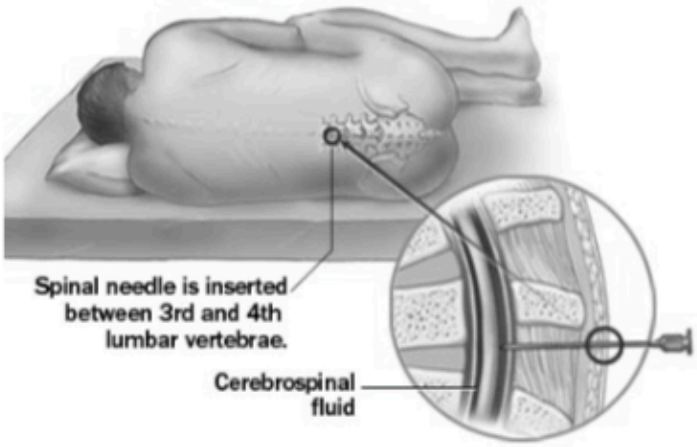
8.2 It is the responsibility of the clinician performing the LP to ensure he or she has had adequate training in the procedure and is supervised when necessary.

8.3 It is the responsibility of the clinician performing the LP to ensure all samples are taken to the

laboratory as soon as the procedure is completed

4. Procedure

4.1 Obtain verbal consent for the lumbar puncture explaining the risks and benefits
4.2 Complete the CSF microbiology request form for the MLW laboratory
4.3 Place the patient lying on his or her side with the knees tightly flexed against the chest
4.4 Identify the landmarks for L4 using the Iliac crest, and palpate the back for the L3/4 space



4.5 Label 1 universal container for CSF with the patient's details. Wash your hands thoroughly and prepare a sterile field.
4.6 Draw up 2ml of 1% lignocaine if available
4.7 Assemble the spinal manometer and ensure the tap is in the correct position to allow CSF to enter the manometer if available
4.8 Clean the skin using alcohol/betadine/chorhexadine skin prep
4.9 Anaesthetise the skin using a small volume of lignocaine, taking care not to obliterate the spinal landmarks for the L3/4 space
4.10 Insert the LP needle (or green cannula if using) into the space carefully at 90 degrees to the skin
4.11 Once the needle has passed through the ligamentum flavum (felt as resistance) withdraw the stylet and when CSF flows attach the manomenter to measure the CSF opening pressure.
4.12 Drain the CSF in the manometer into the first universal container. Continue to allow CSF to drip into this container until >2 mls is obtained (ideally 5mls). Close the lid.
4.13 Remove the needle from the patient and place a small sterile dressing over the wound.
4.14 Inform the patient that he or she can move around after the LP, but a headache is to be expected in some cases. If the headache is bad the patient can have paracetamol. Bed rest is NOT required after a lumbar puncture and will not minimize headache. The patient can be advised to drink a caffeinated drink such as coke to increase CSF production and minimize headache.
4.15 If there are any concerns or complications you must call the senior clinician on duty as soon as is feasible.
4.16 If CSF is not obtained from the L3/4 space then the L4/5 space may be used. The patient can also be asked to sit up, with the back flexed forward to facilitate lumbar puncture from either of these spaces. Please note that CSF opening pressure cannot be recorded from this position.
4.17 When the procedure is completed, document in the patient notes your name, grade and signature, the number of passes made to obtain CSF, the volume of local anaesthetic (if used), and the appearance and opening pressure of the CSF in the notes.

4.2 Disposal of waste and sharps

4.3.1 All sharps should be disposed of in a designated sharps container if available
4.3.2 Do not dispose of sharps in a standard bin
4.3.3 Do not re-sheath sharps

4.4 All clinical waste should be disposed of in the appropriate place in the AETC

5. Health and Safety

5.1 CSF is a body fluid that can become infected with blood borne viruses and as such poses a risk in the event of a needle stick injury or CSF splash, particularly of HIV infection. If such a needlestick event occurs, you must notify your line manager and follow the procedure for a blood needle stick

10.3 Per-Nasal swab SOP

F.2 Per-nasal swab

This is taken by passing a pernasal swab along the floor of the nose until it reaches the posterior wall of the naso-pharynx, left for a maximum of 5 seconds and withdrawn and placed back into the transport media[F1]. Gentle upward pressure on the tip of the nose will expose the nostril and anterior nose. Label with study number, date and name and send to laboratory. The swab should pass without resistance until it reaches the posterior pharynx, which is approximately ½ the distance from the nostril to the ear lobe. If resistance is encountered, the swab should be removed and an attempt made to pass the swab through the other nostril. Often, failure to obtain a good specimen results from the failure to pass the swab fully into the posterior nasopharynx. A record should be kept of the presence or absence of nasal mucous and the success of the procedure as acceptable or suboptimal.

Once taken the swab should be placed into the supplied vials of skim milk-tryptone-glucose-glycerin (STGG) (Appendix H 3.1)by cutting the swab with scissors – scissors will need to be flamed using a cigarette lighter. If this cannot be done the swab should be placed back into the swab sheath and delivered to the laboratory for immersion in the STGG as soon as possible after which plating can be performed.

Pernasal swabs

Pernasal swabs will be delivered to the laboratory as a complete swab or as a tip in skim milk-tryptone-glucose-glycerin (STGG) transport medium. The STGG transport media production is based on the formula described by Gibson and Khoury[G1,G2].

If the laboratory receive a complete swab the tip should be cut off into a vial of STGG media and left to stand for approximately 15 minutes after which the sample can be treated as if it had been delivered already in STGG.

The vial of STGG should be vortexed and a 10µL loop used to inoculate a Columbia blood agar (BA) plate supplemented with blood from horse, sheep, or goat. Selective BA plates with 2.5 mcg/mL or 5.0 mcg/mL of gentamicin (GBA), colistin/nalidixic acid or colistin-oxolinic acid are considered as part of the core WHO method and may be required if growth of contaminant from pernasal swabs proves heavy. Incubation of the plate should take place at 35°C in 3% to 10% CO_2 overnight.

From the primary plate pick and streak out on one half of a BA plate two presumptive pneumococcal colonies, attempting to select colonies which look as different as possible; place an optochin disc in the centre of each streak, incubate overnight as above.

Susceptibility to optochin is defined as the diameter of inhibition zone. Zones of inhibition greater than 14 mm indicate susceptibility, 7-13 mm are indeterminate and less than 7 mm is resistant to optochin when a 6 mm size disc is used and the culture is incubated in 5% CO_2. Isolates, which are optochin susceptible, are considered pneumococci, those of intermediate susceptibility

---

[G1] O'Brien KL, Bronsdon MA, Dagan R, et al. Evaluation of a medium (STGG) for transport and optimal recovery of Streptococcus pneumoniae from nasopharyngeal secretions collected during field studies. J Clin Microbiol 2001; 39(3):1021-1024.

[G2] Gibson LF, Khoury JT. Storage and survival of bacteria by ultra-freeze. Letters in Applied Microbiology 1986; 3:127-129.

should be tested for bile solubility. Those that are optochin resistant are considered to be a species other than pneumococcus, however, a small number of pneumococci may be optochin resistant. Serotype and store pneumococcal strains as noted below.

The vial containing the per-nasal tip should be frozen at –70°C following inoculation of the plate.

10.4 H1N1, Nasopharnygeal Aspirate Sampling

H1N1 Screening Time Point 0 (T_0) – Clinical Diagnostic Sampling SOP

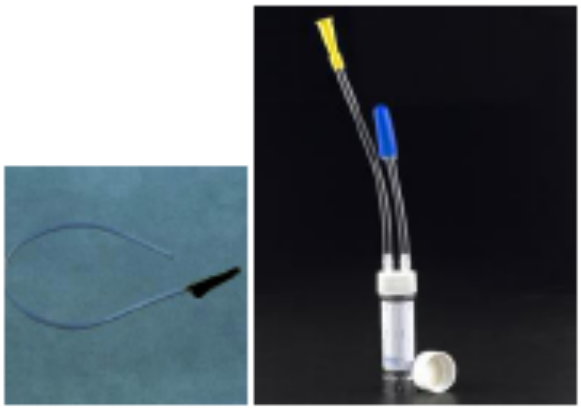
Important

- Verbal (as opposed to written) consent is required; this can be stated to either the clinician or the H1N1 researcher
- Consent to undergo clinical diagnostic sampling does not represent consent to participate in the H1N1 study, even when the sampling is performed by a H1N1 researcher
- The T_0 diagnostic sample is ideally taken by a trained H1N1 researcher to encourage standardisation in sampling

Sampling methods

Nasopharyngeal aspirate sampling is the method of choice.

Equipment



- 10 CH black-tip Argyle suction catheter
- Suction trap pack with container lid
- Connector tubing to connect suction trap to wall suction
- 5 ml plastic vial of sterile 0.9% saline
- 5 ml disposable plastic syringe
- Disposable Pasteur pipette
- Sterile universal container

PPE and infection control requirements
According to QECH guidelines. At the time of writing, the operator should wear:

- disposable gloves
- plastic gown
- face mask
- eye protection

Method

1. Explain to the patient/child and/or parent/guardian that NPA is not a painful sampling procedure but that it can cause some mild nasal irritation (e.g. tickling sensation, feeling of wanting to sneeze and eye watering).

2. Using gloved hands, remove the suction trap from its packaging and attach the suction trap to suction canister of the portable suction using an appropriate length of plastic connector tubing connected to the suction trap port. Keep the suction trap screw-top lid in the packaging

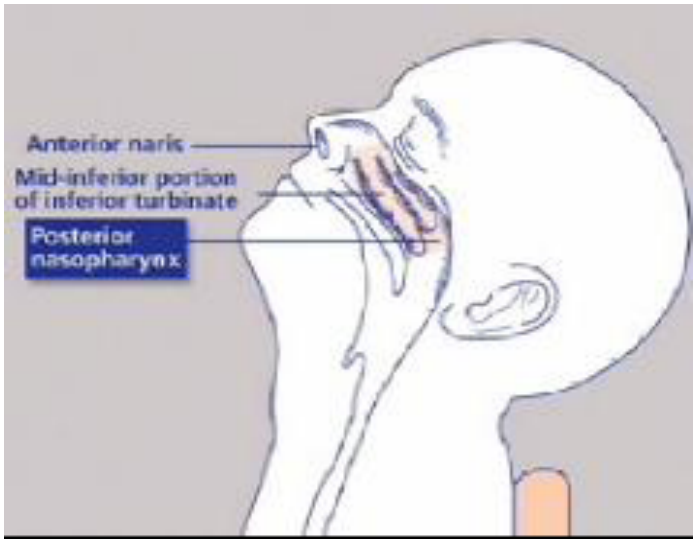
3. Open the end of the suction catheter tubing nearest to the black suction control on the catheter, keeping the catheter within its packaging until ready for use. Attach the black end of the catheter to the remaining port on the suction trap.

4. Remove the top of the vial of the sterile saline and place upright on a surface within easy reach. Remove and discard 2ml normal saline using a disposable sterile 5 ml syringe.

5. With the patient either sitting upright or with on a trolley with the back positioned at around 45°, tilt the patient's head back slightly

6. Switch on wall suction to **13 kPa** (100 mmHg) and maintaining an aseptic technique, remove the catheter from its packaging

7. Without applying suction, insert the catheter into the patient's **LEFT nostril** to reach the posterior nasopharynx, approximately half way along an imaginary line drawn between the left nostril and the left ear lobe



8. Apply suction by occluding the hole in the black end of the suction catheter and then withdraw the catheter over 5 seconds whilst rotating your wrist. **Keep the suction trap upright during this procedure**. Release suction when the catheter has been removed from the patient's nose.

9. Examine the catheter and trap to look for secretions coating the inside of the catheter and trap. If no secretions are seen, attempt to repeat the procedure.

10. Place the tip of the catheter in the vial of saline, apply suction and aspirate the entire contents (3ml) into the trap. Release suction and turn off the wall suction.

11. Unscrew the top (port lid) of the suction trap, remove the screw-cap from the packaging and screw onto the sputum trap container.

12. Dispose of the catheter, suction tubing and sputum trap port lid in the clinical waste bin.

13. Label the pot with the appropriate **H1N1 SPINE patient identifying details/pre-printed SPINE label** and place in a specimen bag.

14. Deliver the specimen to the diagnostic laboratory reception.

15. Record the patient's details and current/planned location within the hospital.

16. Document in the patient's clinical notes the time and date of sampling.

Alternative sampling procedure for those unable to tolerate aspirate collection:

Flocked nasopharyngeal swab

Sampling Procedure: Flocked nasopharyngeal swab sampling

1. Inform the patient and/or parent/guardian that you are going to perform NP swab sampling by inserting **3** small swabs into the back of the nose.

2. Inform the patient and/or parent/guardian that although uncomfortable, the procedure should not be painful. The most common side effect is nasal irritation and watering eyes; both side-effects are very short lived.

3. Sit the patient upright with their head tilted back slightly, or if on a trolley/bed, sit the patient at 45 degrees with their head resting on the trolley/bed.

4. Equipment: **Flocked nylon nasopharyngeal swab, 2x 1ml pot of universal transport medium (UTM) and one cryotube of STGG**. Check expiry date prior to use.

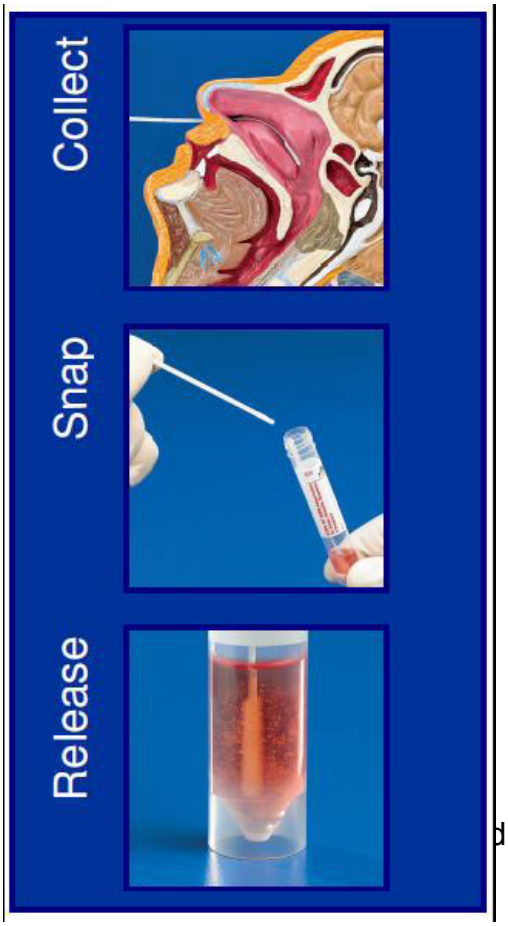
Note that the UTM may have a green screw cap

5. Wearing personal protective equipment (PPE), remove the swab from its packaging, instruct the patient to close their eyes (to minimise discomfort) and gently insert the swab into the **LEFT nostril**, advancing the swab along the floor of the nasal passage until the resistance is felt at the posterior wall of the nasopharynx. This should

be approximately half the length of an imaginary line drawn from ear

lobe.

6. Rotate the swab 360° three times and then remove the swab, placing it in the pot of universal transport medium (UTM), snapping off the tip within the pot.

7. 

d discard the swab handle in a

8. Label the pot with the appropriate **H1N1 SPINE patient identifying details/pre-printed SPINE label** and place in a specimen bag **for delivery to MLW EVERY hour**

9. Repeat steps 5-8 remembering to placing the second swab into UTM and the third swab into the STGG cryotube.

10. Deliver the specimen to the diagnostic laboratory reception.

11. Record the patient's details and current/planned location within the hospital.

12. Document in the patient's clinical notes the time and date of sampling.

10.5 STGG media SOP

H.3.1.1 STGG media preparation

Mix together the following products. The manufacturer and lot numbers of ingredients should be standardised for a given study whenever possible.

Oxoid tryptone soya broth (CM 129)	3.0 g
Glucose	0.5 g
Oxoid skim milk powder (CM L31)	2.0 g
Glycerol	10.0 mL
Double distilled water	100.0 mL

Dispense in 0.5 to 1.0 mL amounts in Nunc or Nalgene cryotubes and autoclave at 15 lb for no more than 10 minutes. Tighten the lids before storage. The tubes may then be stored refrigerated (4-6 °C) or at room temperature (22 °C) for several months. Prior to use, the pellet in the bottom of the tube should be resuspended by vortexing for at least 10-15 seconds.

Quality control of STGG transport medium should be conducted to assure sterility and the ability of the media to sustain pneumococci. The solution should be tan in colour with a precipitate at the bottom of the tube.

Figure 81: Excerpt from the PNEUVAC study protocol and procedure manual

10.6 1x Phosphate Buffered Saline

1. Dissolve the following in 800ml distilled H_2O.
    - 8g of NaCl
    - 0.2g of KCl
    - 1.44g of Na_2HPO_4
    - 0.24g of KH_2PO_4
    -
2. Adjust pH to 7.4 with HCl.

3. Adjust volume to 1L with additional distilled H_2O.
4. Sterilize by autoclaving.

10.7 CSP sequence

CSP (0100993)	ATGAAAAACACAGTTAAATTGGAACAGTTTGTAG CTTTGAAGGAAAAAGACTTACAAAAGATTAAAGG TGGGGAGATGAGGCTGTCAAAATTCTTCCGTGAT TTTATTTTACAAAGAAAAAAGTAA

Figure 82: Competence stimulating peptide used for experiments with isolate 0100993

10.8 Amino acid abbreviation table

Amino acid	Abbreviation
Alanine	Ala/A
Arginine	Arg/R
Asparagine	Asn/N
Aspartic acid	Asp/D
Cysteine	Cys/C
Glutamic acid	Glu/E
Glutamine	Gln/Q
Glycine	Gly/G
Histidine	His/H
Isoleucine	Ile/I
Leucine	Leu/L
Lysine	Lys/K
Methionine	Met/M
Phenylalanine	Phe/F
Proline	Pro/P
Serine	Ser/S
Threonine	Thr/T
Tryptophan	Trp/W
Tyrosine	Tyr/Y
Valine	Val/V

10.9- GWAS significant SNPs

Gene	Amox	Amp	Pe	Cefo	Cefta	Ceftri	Unique
alsS	1	1					1
aroA	2						2
aroB	2		7				9
aroF			4				4
asnA	1						1
cbpD	3		1				4
clpC	1						1
ddlA	10	4					10
ddlA promoter	8	8					8
fsaA	4	6					6
ftsH			1				1
ftsL	1						1
fus	1	1					1
gldA	3	4					4
gpmA			1				1
guaA		1	1				1
htrA			1				1
ileS			9				9
leuS	21	21					21
mraW	1						1
murF	11	4					11
nanB		3	3				3
pbp2B	12	10					14
pbp2X	1			2	4		7
pflD	3	3					3
plr			1				1
plsX			3				3
pth					1		1
purF	1						1
recR	5	2					5
scaR	2						2
SPN23F00040	1						1
SPN23F00080	1						1
SPN23F00530			1				1
SPN23F00790	4						4
SPN23F01410	1						1
SPN23F01440			1				1

Gene	Amox	Amp	Pe	Cefo	Cefta	Ceftri	Unique
SPN23F01460	3						3
SPN23F01471	1						1
SPN23F01490			1				1
SPN23F01510			1				1
SPN23F01520	5						5
SPN23F01530	1		5				6
SPN23F01540			8				8
SPN23F03400						1	1
SPN23F05500		1					1
SPN23F05610	1						1
SPN23F06970	1						1
SPN23F07110		1					1
SPN23F07240	1		3				4
SPN23F08780			1				1
SPN23F10060			1				1
SPN23F10370	5	6					6
SPN23F11350	1						1
SPN23F11460	1						1
SPN23F13380			1				1
SPN23F14110		1	1				1
SPN23F14120			2				2
SPN23F15670	1						1
SPN23F16540	1						
SPN23F16560			7				7
SPN23F16590			4				4
SPN23F20320			2				2
SPN23F20340	1						1

Gene	Amox	Amp	Pe	Cefo	Cefta	Ceftri	Unique
SPN23F20350	1						1
SPN23F20380	1						1
SPN23F20630	1						1
SPN23F21180	1						1
SPN23F21520			1				1
SPN23F21620			1				1
SPN23F22260	1						1
SPN23F22380	2		1				3
SPN23F22390	1						1
T-box leader			1				1
thrS			1				1
treA			1				1
tyrA	1						1
intergenic	31	12	17				48

10.10 Insertion size distribution

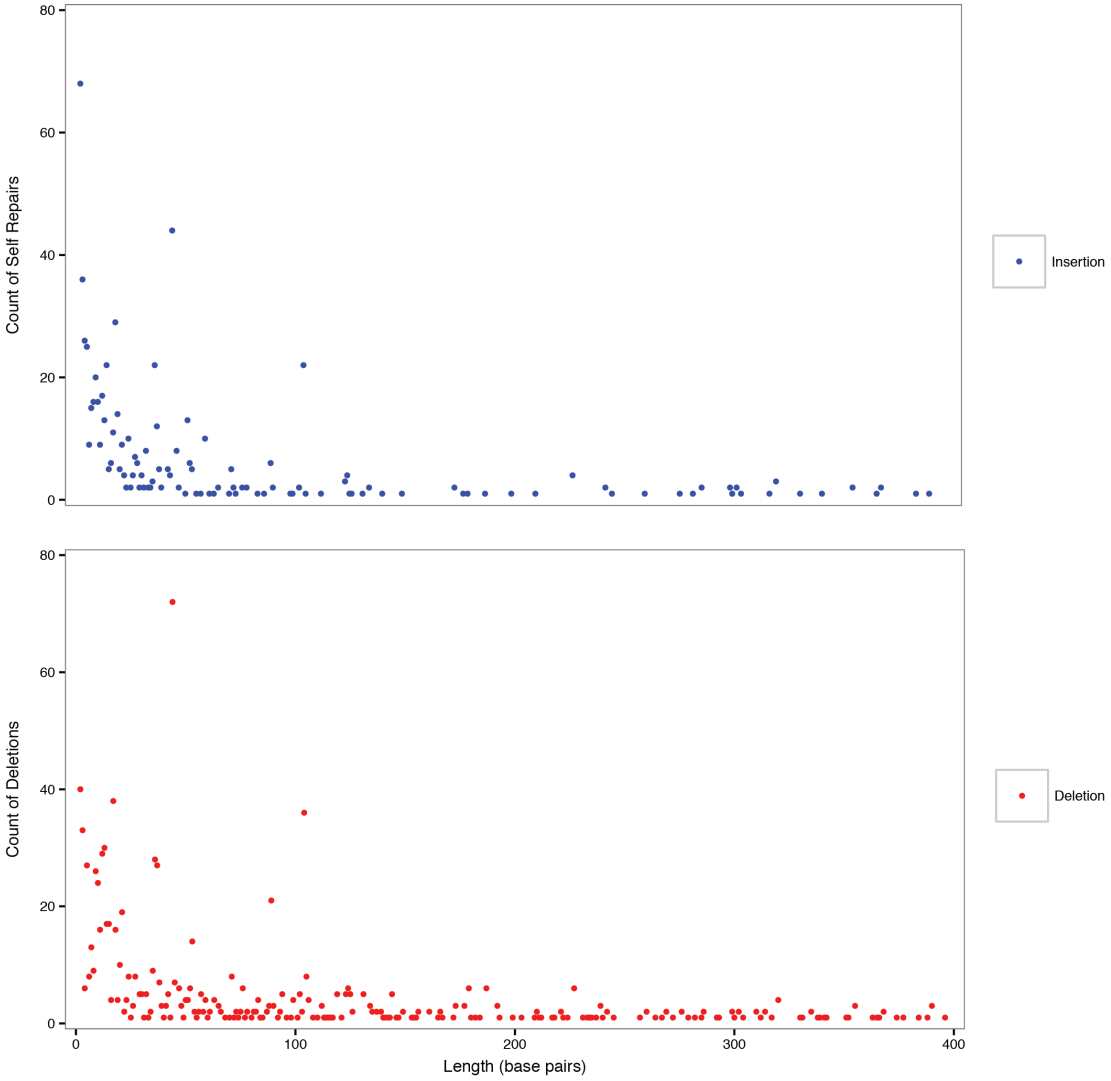


Figure 83: Size distributions of repairs and deletions.

10.11 Raw simulation data